



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/00	A2	(11) International Publication Number: WO 98/56902 (43) International Publication Date: 17 December 1998 (17.12.98)
(21) International Application Number: PCT/US98/11808 (22) International Filing Date: 9 June 1998 (09.06.98) (30) Priority Data: 871,755 9 June 1997 (09.06.97) US (71) Applicant (for all designated States except US): MEDICAL COLLEGE OF GEORGIA RESEARCH INSTITUTE, INC. [US/US]; 1120 15th Street, Augusta, GA 30912-4810 (US). (72) Inventor; and (75) Inventor/Applicant (for US only): LIN, Shuo [-/US]; 1120 15th Street, Augusta, GA 30912-4810 (US). (74) Agents: PABST, Patrea, L. et al.; Amall Golden & Gregory, LLP, 2800 One Atlantic Center, 1201 West Peachtree Street, Atlanta, GA 30309-3450 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: TRANSGENIC FISH WITH TISSUE-SPECIFIC EXPRESSION (57) Abstract <p>Disclosed are transgenic fish, and a method of making transgenic fish, which express transgenes in stable and predictable tissue- or developmentally-specific patterns. The transgenic fish contain transgene constructs with homologous expression sequences. Also disclosed are methods of using such transgenic fish. Such expression of transgenes allows the study of developmental processes, the relationship of cell lineages, the assessment of the effect of specific genes and compounds on the development or maintenance of specific tissues or cell lineages, and the maintenance of lines of fish bearing mutant genes.</p>		

BEST AVAILABLE COPY

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

TRANSGENIC FISH WITH TISSUE-SPECIFIC EXPRESSION

BACKGROUND OF THE INVENTION

5 The disclosed invention is generally in the field of transgenic fish, and more specifically in the area of transgenic fish exhibiting tissue-specific expression of a transgene.

 Transgenic technology has become an important tool for the study of gene and promoter function (Hanahan, *Science* **246**:1265-75 (1989);
10 Jaenisch, *Science* **240**:1468-74 (1988)). The ability to express, and study the expression of, genes in whole animals can be facilitated by the use of transgenic animals. Transgenic technology is also a useful tool for cell lineage analysis and for transplantation experiments. Studies on promoter function or lineage analysis generally require the expression of a foreign
15 reporter gene, such as the bacterial gene *lacZ*. Expression of a reporter gene can allow the identification of tissues harboring a transgene. Typically, transgenic expression has been identified by *in situ* hybridization or by histochemistry in fixed animals. Unfortunately, the inability to easily detect transgene expression in living animals severely
20 limits the utility of this technology, particularly for lineage analysis.

 An attractive paradigm for the understanding of gene expression, development, and genetics of animals, especially humans, is to study less complex organisms, such as *Escherichia coli*, *Drosophila*, and
25 *Caenorhabditis*. The hope is that understanding of these processes in simple organisms will have relevance to similar processes in mammals and humans. The tradeoff is to accept the disadvantage that an experimental organism is only distantly related to humans for the advantage of easy manipulation, fast generation times, and more straightforward interpretation of results in the experimental organism.
30 The disadvantage of this tradeoff can be lessened by using an organism that is as closely related as possible to mammals while retaining as many of the advantages of less complex organisms. The problem is to identify

suitable organisms for such studies, and, more importantly, to develop the tools necessary to manipulate such organisms.

Some examples of cell determination in invertebrates have been shown to occur in progressive waves that are regulated by sequential
5 cascades of transcription factors. Much less is known about such processes in vertebrates. An integrated approach combining embryological, genetic and molecular methods, such as that used to study development in *Drosophila* (for example, Ghysen *et al.*, *Genes & Dev* 7:723-33 (1993)), would facilitate the identification of the molecular
10 mechanisms involved in specifying neuronal fates in vertebrates, but such an approach has been hampered by a lack of robust genetic and molecular tools for use in vertebrates.

Transgenic technology has been applied to fish for various purposes. For example, transgenic technology has been applied to several
15 commercially important varieties of fish, primarily in an attempt to improve their cultivation. The use of transgenic technology in fish has been reviewed by Moav, *Israel J. of Zoology* 40:441-466 (1994), Chen *et al.*, *Zoological Studies* 34:215-234 (1995), and Iyengar *et al.*, *Transgenic Res.* 5:147-166 (1996).

20 Stuart *et al.*, *Development* 103:403-412 (1988), describe integration of foreign DNA into zebrafish, but no expression was observed. Stuart *et al.*, *Development* 109:577-584 (1990), describe expression of a transgene in zebrafish from SV40 and Rous sarcoma virus transcription regulatory sequences. Although expression was seen in a
25 pattern of tissues, the expression within a given tissue was variegated. Also, since Stuart *et al.* (1990) selected transgenics by expression and not by the presence of the transgene, non-expressing transgenics would have been missed by their analysis. Culp *et al.*, *Proc. Natl. Acad. Sci. USA* 88:7953-7957 (1991), describe integration and germ line transmission of
30 DNA in zebrafish. Although the constructs used included the Rous sarcoma virus LTR or SV40 enhancer promoter linked to a *lacZ* gene, no expression was observed. Bayer and Campos-Ortega, *Development*

115:421-426 (1992), describe integration and expression in zebrafish of a *lacZ* transgene having a minimal promoter (a mouse heat shock promoter) but no upstream regulatory sequences. The expression obtained depended on the site of integration indicating that endogenous sequences at the site of integration of the fish were responsible for expression. Westerfield *et al.*, *Genes & Development* 6:591-598 (1992), describe transient expression in zebrafish of β -galactosidase from mouse and human *Hox* gene promoters. Lin *et al.*, *Dev. Biology* 161:77-83 (1994), describe transgenic expression of *lacZ* in living zebrafish embryos. The transgene linked the enhancer-promoter of the *Xenopus* elongation factor 1 α gene with the *lacZ* coding sequence. Different lines of transgenic fish exhibited different patterns of expression, indicating that the site of integration may be affecting the pattern of expression. Amsterdam *et al.*, *Dev. Biology* 171:123-129 (1995), and Amsterdam *et al.*, *Gene* 173:99-103 (1996), describe transgenic expression of green fluorescent protein (GFP) in zebrafish. The transgene linked the enhancer-promoter of the *Xenopus* elongation factor 1 α gene with the GFP coding sequence. As in Lin *et al.*, *Dev. Biology* 161:77-83 (1994), different lines of transgenic fish exhibited different patterns of expression, indicating that the site of integration may be affecting the pattern of expression. Although some of the systems described above exhibited patterned expression, none resulted in the transmission of stable tissue-specific expression of a transgene in zebrafish.

It is an object of the present invention to provide transgenic fish having tissue- and developmentally-specific expression of transgenes.

It is another object of the present invention to provide a method of making transgenic fish having tissue- and developmentally-specific expression of transgenes.

It is another object of the present invention to provide a method of identifying compounds that affect expression of fish genes of interest.

It is another object of the present invention to provide a method of identifying the pattern of expression of fish genes of interest.

It is another object of the present invention to provide a method of identifying genes that affect expression of fish genes of interest.

It is another object of the present invention to provide a method of genetically marking mutant fish genes.

5 It is another object of the present invention to provide a method of identifying fish that have inherited a mutant gene.

It is another object of the present invention to provide a method of identifying enhancers and other regulatory sequences in fish.

10 It is another object of the present invention to provide a construct that exhibits tissue- and developmentally-specific expression in fish.

BRIEF SUMMARY OF THE INVENTION

Disclosed are transgenic fish, and a method of making transgenic fish, which express transgenes in stable and predictable tissue- or
15 developmentally-specific patterns. The transgenic fish contain transgene constructs with homologous expression sequences. Also disclosed are methods of using such transgenic fish. Such expression of transgenes allow the study of developmental processes, the relationship of cell
20 lineages, the assessment of the effect of specific genes and compounds on the development or maintenance of specific tissues or cell lineages, and the maintenance of lines of fish bearing mutant genes.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A shows the nucleotide sequence at the exon/intron
25 junctions of the zebrafish GATA-1 locus. The conserved splice sequences are underlined and the intron sequences are listed within parentheses. The amino acids encoded by the exon regions flanking the introns are shown beneath the nucleotide sequence. The upstream splice junction nucleotide sequences are SEQ ID NO:6 (IVS-1), SEQ ID NO:7 (IVS-2),
30 SEQ ID NO:8 (IVS-3), and SEQ ID NO:9 (IVS-4). The downstream splice junction nucleotide sequences are SEQ ID NO:10 (IVS-1), SEQ ID NO:11 (IVS-2), SEQ ID NO:12 (IVS-3), and SEQ ID NO:13 (IVS-4).

The amino acid sequences spanning the introns are SEQ ID NO:14 (IVS-1), SEQ ID NO: 15 (IVS-2), SEQ ID NO:16 (IVS-3), and SEQ ID NO:17 (IVS-4).

Figure 1B is a diagram of the structure of the zebrafish GATA-1 locus. Exon regions are filled. Intron regions are unfilled. The tall filled boxes represent the coding regions. The arrow indicates the putative transcription start site. *EcoRI* endonuclease sites are labeled E. *BglII* endonuclease sites are labeled G. *BamHI* endonuclease sites are labeled B.

Figure 2 is a diagram of the structures of three GATA-1/GFP transgene constructs used to make transgenic fish. The filled region to the right of the GM2 box in each construct represents the 5.4 kb or 5.6 kb region of the GATA-1 locus upstream of the GATA-1 coding region. The box labeled GM2 represents a sequence encoding the modified green fluorescent protein. The thin angled lines in constructs (1) and (3) represent vector or linking sequences. *EcoRI* endonuclease sites are labeled E. *BglII* endonuclease sites are labeled G. *BamHI* endonuclease sites are labeled B. In construct (3), the *BamHI/EcoRI* fragment on the right side is the downstream *BamHI/EcoRI* fragment of the GATA-1 locus.

Figure 3 is a diagram of the structures of GATA-2/GFP transgene constructs for analyzing the expression sequences of the GATA-2 gene. The line represents all or upstream deleted portions of a 7.3 kb region upstream of the translation start site in the zebrafish GATA-2 gene. The hatched box represents a segment encoding the modified GFP and including a SV40 polyadenylation signal. Tick marks labeled P, Sa, A, C, and Sc indicates restriction sites *PstI*, *SacI*, *AatII*, *ClaI* and *ScaI*, respectively, in the 7.3 kb region.

Figure 4 is a diagram of the structures of GATA-2/GFP transgene constructs for analyzing the expression sequences of the GATA-2 gene. The thick open box represents a 1116 bp fragment of the upstream region of the GATA-2 gene required for neuron-specific expression. The thin

open box represents segments of the upstream region of the GATA-2 gene proximal to the transcription start site. The thick line represents the minimal promoter of the *Xenopus* elongation factor 1 α gene. The hatched box represents a segment encoding the modified GFP and including a SV40 polyadenylation signal.

Figure 5 is a graph of the percent of embryos microinjected with the transgene constructs shown in Figure 4 that expressed GFP in neurons.

Figure 6 is a graph of the percent of embryos microinjected with transgene constructs that expressed GFP in neurons. The transgene constructs were nsP5-GM2 and truncated forms of nsP5-GM2.

Figure 7 is a graph of the percent of embryos microinjected with transgene constructs that expressed GFP in neurons. The transgene constructs were mutant forms of the ns3831 truncation of nsP5-GM2.

DETAILED DESCRIPTION OF THE INVENTION

Disclosed are transgenic fish, and a method of making transgenic fish, which express transgenes in stable and predictable tissue- or developmentally-specific patterns. Also disclosed are methods of using such transgenic fish. Such expression of transgenes allow the study of developmental processes, the relationship of cell lineages, the assessment of the effect of specific genes and compounds on the development or maintenance of specific tissues or cell lineages, and the maintenance of lines of fish bearing mutant genes. The disclosed transgenic fish are characterized by homologous expression sequences in an exogenous construct introduced into the fish or a progenitor of the fish.

As used herein, transgenic fish refers to fish, or progeny of a fish, into which an exogenous construct has been introduced. A fish into which a construct has been introduced includes fish which have developed from embryonic cells into which the construct has been introduced. As used herein, an exogenous construct is a nucleic acid that is artificially introduced, or was originally artificially introduced, into an animal. The

term artificial introduction is intended to exclude introduction of a construct through normal reproduction or genetic crosses. That is, the original introduction of a gene or trait into a line or strain of animal by cross breeding is intended to be excluded. However, fish produced by transfer, through normal breeding, of an exogenous construct (that is, a construct that was originally artificially introduced) from a fish containing the construct are considered to contain an exogenous construct. Such fish are progeny of fish into which the exogenous construct has been introduced. As used herein, progeny of a fish are any fish which are descended from the fish by sexual reproduction or cloning, and from which genetic material has been inherited. In this context, cloning refers to production of a genetically identical fish from DNA, a cell, or cells of the fish. The fish from which another fish is descended is referred to as a progenitor fish. As used herein, development of a fish from a cell or cells (embryonic cells, for example), or development of a cell or cells into a fish, refers to the developmental process by which fertilized egg cells or embryonic cells (and their progeny) grow, divide, and differentiate to form an adult fish.

The examples illustrate the manner in which transgenic fish exhibiting cell lineage-specific expression can be made and used. The transgenic fish described in the examples, and the transgene constructs used, are particularly useful for early detection of fish expressing the transgene, the study of erythroid cell development, the study of neuronal development, and as a reporter for genetically linked mutant genes.

Tissue-, developmental stage-, or cell lineage-specific expression of a reporter gene from a regulated promoter in the disclosed transgenic fish can be useful for identifying the pattern of expression of the gene from which the promoter is derived. Such expression can also allow study of the pattern of development of a cell lineage. As used herein, tissue-specific expression refers to expression substantially limited to specific tissue types. Tissue-specific expression is not necessarily limited to expression in a single tissue but includes expression limited to one or

more specific tissues. As used herein, developmental stage-specific expression refers to expression substantially limited to specific developmental stages. Developmental stage-specific expression is not necessarily limited to expression at a single developmental stage but
5 includes expression limited to one or more specific developmental stage. As used herein, cell lineage-specific expression refers to expression substantially limited to specific cell lineages. As used herein, cell lineage refers to a group of cells that are descended from a particular cell or group of cells. In development, for example, newly specialized or
10 differentiated cells can give rise to cell lineages. Cell lineage-specific expression is not necessarily limited to expression in a single cell lineage but includes expression limited to one or more specific cell lineages. All of these types of specific expression can operate in the same gene. For example, a developmentally regulated gene can be expressed at both
15 specific developmental stages and be limited to specific tissues. As used herein, the pattern of expression of a gene refers to the tissues, developmental stages, cell lineages, or combinations of these in or at which the gene is expressed.

1. Transgene Constructs

20 Transgene constructs are the genetic material that is introduced into fish to produce a transgenic fish. Such constructs are artificially introduced into fish. The manner of introduction, and, often, the structure of a transgene construct, render such a transgene construct an exogenous construct. Although a transgene construct can be made up of
25 any nucleic acid sequences, for use in the disclosed transgenic fish it is preferred that the transgene constructs combine expression sequences operably linked to a sequence encoding an expression product. The transgenic construct will also preferably include other components that aid expression, stability or integration of the construct into the genome of a
30 fish. As used herein, components of a transgene construct referred to as being operably linked or operatively linked refer to components being so connected as to allow them to function together for their intended

purpose. For example, a promoter and a coding region are operably linked if the promoter can function to result in transcription of the coding region.

A. Expression Sequences

5 Expression sequences are used in the disclosed transgene constructs to mediate expression of an expression product encoded by the construct. As used herein, expression sequences include promoters, upstream elements, enhancers, and response elements. It is preferred that the expression sequences used in the disclosed constructs be homologous
10 expression sequences. As used herein, in reference to components of transgene constructs used in the disclosed transgenic fish, homologous indicates that the component is native to or derived from the species or type of fish involved. Conversely, heterologous indicates that the component is neither native to nor derived from the species or type of fish
15 involved.

Two large scale chemical mutagenesis screens recently produced thousands of zebrafish mutants affecting development (Driever *et al.*, *Development* 123:37-46 (1996); Haffter *et al.*, *Development* 123:1-36 (1996)). Such genes and their expression patterns are of significant
20 interest for understanding the developmental process. Therefore, expression sequences from these genes are preferred for use as expression sequences in the disclosed constructs.

As used herein, expression sequences are divided into two main classes, promoters and enhancers. A promoter is generally a sequence or
25 sequences of DNA that function when in a relatively fixed location in regard to the transcription start site. A promoter contains core elements required for basic interaction of RNA polymerase and transcription factors, and may contain upstream elements and response elements. Enhancer generally refers to a sequence of DNA that functions at no fixed
30 distance from the transcription start site and can be in either orientation. Enhancers function to increase transcription from nearby promoters. Enhancers also often contain response elements that mediate the regulation

of transcription. Promoters can also contain response elements that mediate the regulation of transcription.

Enhancers often determine the regulation of expression of a gene. This effect has been seen in so-called enhancer trap constructs where
5 introduction of a construct containing a reporter gene operably linked to a promoter is expressed only when the construct inserts into the domain of an enhancer (O'Kane and Gehring, *Proc. Natl. Acad. Sci. USA* **84**:9123-9127 (1987), Allen *et al.*, *Nature* **333**:852-855 (1988), Kothary *et al.*, *Nature* **335**:435-437 (1988), Gossler *et al.*, *Science* **244**:463-465 (1989)).
10 In such cases, the expression of the construct is regulated according to the pattern of the newly associated enhancer. Transgenic constructs having only a minimal promoter can be used in the disclosed transgenic fish to identify enhancers.

Preferred enhancers for use in the disclosed transgenic fish are
15 those that mediate tissue- or cell lineage-specific expression. More preferred are homologous enhancers that mediate tissue- or cell lineage-specific expression. Still more preferred are enhancers from fish GATA-1 and GATA-2 genes. Most preferred are enhancers from zebrafish GATA-1 and GATA-2 genes.

20 For expression of encoded peptides or proteins, a transgene construct also needs sequences that, when transcribed into RNA, mediate translation of the encoded expression products. Such sequences are generally found in the 5' untranslated region of transcribed RNA. This region corresponds to the region on the construct between the
25 transcription initiation site and the translation initiation site (that is, the initiation codon). The 5' untranslated region of a construct can be derived from the 5' untranslated region normally associated with the promoter used in the construct, the 5' untranslated region normally associated with the sequence encoding the expression product, the 5'
30 untranslated region of a gene unrelated to the promoter or sequence encoding the expression product, or a hybrid of these 5' untranslated regions. Preferably, the 5' untranslated region is homologous to the fish

into which the construct is to be introduced. Preferred 5' untranslated regions are those normally associated with the promoter used.

B. Expression Products

Transgene constructs for use in the disclosed transgenic fish can
5 encode any desired expression product, including peptides, proteins, and
RNA. Expression products can include reporter proteins (for detection
and quantitation of expression), and products having a biological effect on
cells in which they are expressed (by, for example, adding a new
enzymatic activity to the cell, or preventing expression of a gene). Many
10 such expression products are known or can be identified.

Reporter Proteins

As used herein, a reporter protein is any protein that can be
specifically detected when expressed. Reporter proteins are useful for
detecting or quantitating expression from expression sequences. For
15 example, operatively linking nucleotide sequence encoding a reporter
protein to a tissue specific expression sequences allows one to carefully
study lineage development. In such studies, the reporter protein serves as
a marker for monitoring developmental processes, such as cell migration.
Many reporter proteins are known and have been used for similar
20 purposes in other organisms. These include enzymes, such as β -
galactosidase, luciferase, and alkaline phosphatase, that can produce
specific detectable products, and proteins that can be directly detected.
Virtually any protein can be directly detected by using, for example,
specific antibodies to the protein. A preferred reporter protein that can be
25 directly detected is the green fluorescent protein (GFP). GFP, from the
jellyfish *Aequorea victoria*, produces fluorescence upon exposure to
ultraviolet light without the addition of a substrate (Chalfie *et al.*, *Science*
263:802-5 (1994)). Recently, a number of modified GFPs have been
created that generate as much as 50-fold greater fluorescence than does
30 wild type GFP under standard conditions (Cormack *et al.*, *Gene* **173**:33-8
(1996); Zolotukhin *et al.*, *J. Virol* **70**:4646-54 (1996)). This level of

fluorescence allows the detection of low levels of tissue specific expression in a living transgenic animal.

The use of reporter proteins that, like GFP, are directly detectable without requiring the addition of exogenous factors are preferred for detecting or assessing gene expression during zebrafish embryonic development. A transgenic zebrafish embryo, carrying a construct encoding a reporter protein and a tissue-specific expression sequences, can provide a rapid real time *in vivo* system for analyzing spatial and temporal expression patterns of developmentally regulated genes.

10 C. Other Construct Sequences

The disclosed transgene constructs preferably include other sequences which improve expression from, or stability of, the construct. For example, including a polyadenylation signal on the constructs encoding a protein ensures that transcripts from the transgene will be processed and transported as mRNA. The identification and use of polyadenylation signals in expression constructs is well established. It is preferred that homologous polyadenylation signals be used in the transgene constructs.

It is also known that the presence of introns in primary transcripts can increase expression, possibly by causing the transcript to enter the processing and transport system for mRNA. It is preferred that an intron, if used, be included in the 5' untranslated region or the 3' untranslated region of the transgene transcript. It is also preferred that the intron be homologous to the fish used, and more preferably homologous to the expression sequences used (that is, that the intron be from the same gene that some or all of the expression sequences are from). The use and importance of these and other components useful for transgene constructs are discussed in Palmiter *et al.*, *Proc. Natl. Acad. Sci. USA* 88:478-482 (1991); Sippel *et al.*, "The Regulatory Domain Organization of Eukaryotic Genomes: Implications For Stable Gene Transfer" in *Transgenic Animals* (Grosveld and Kollias, eds., Academic Press, 1992), pages 1-26; Kollias and Grosveld, "The Study of Gene Regulation in

Transgenic Mice" in *Transgenic Animals* (Grosveld and Kollias, eds, Academic Press, 1992), pages 79-98; and Clark *et al.*, *Phil. Trans. R. Soc. Lond. B.* 339:225-232 (1993).

5 The disclosed constructs are preferably integrated into the genome of the fish. However, the disclosed transgene construct can also be constructed as an artificial chromosome. Such artificial chromosomes containing more than 200 kb have been used in several organisms. Artificial chromosomes can be used to introduce very large transgene constructs into fish. This technology is useful since it can allow faithful
10 recapitulation of the expression pattern of genes that have regulatory elements that lie many kilobases from coding sequences.

2. Fish

The disclosed constructs and methods can be used with any type of fish. As used herein, fish refers to any member of the classes
15 collectively referred to as pisces. It is preferred that fish belonging to species and varieties of fish of commercial or scientific interest be used. Such fish include salmon, trout, tuna, halibut, catfish, zebrafish, medaka, carp, tilapia, goldfish, and loach.

The most preferred fish for use with the disclosed constructs and
20 methods is zebrafish, *Danio rerio*. Zebrafish are an increasingly popular experimental animal since they have many of the advantages of popular invertebrate experimental organisms, and include the additional advantage that they are vertebrates. Another significant advantage of zebrafish for the study of development and cell lineages is that, like *Caenorhabditis*,
25 they are largely transparent (Kimmel, *Trends Genet* 5:283-8 (1989)). The generation of thousands of zebrafish mutants (Driever *et al.*, *Development* 123:37-46 (1996); Haffter *et al.*, *Development* 123:1-36 (1996)) provides abundant raw material for transgenic study of these animals. General zebrafish care and maintenance is described by Streisinger, *Natl. Cancer*
30 *Inst. Monogr.* 65:53-58 (1984).

Zebrafish embryos are easily accessible and nearly transparent. Given these characteristics, a transgenic zebrafish embryo, carrying a

construct encoding a reporter protein and tissue-specific expression sequences, can provide a rapid real time *in vivo* system for analyzing spatial and temporal expression patterns of developmentally regulated genes. In addition, embryonic development of the zebrafish is extremely rapid. In 24 hours an embryo develops rudiments of all the major organs, including a functional heart and circulating blood cells (Kimmel, *Trends Genet* 5:283-8 (1989)). Other fish with some or all of the same desirable characteristics are also preferred.

3. Production of Transgenic Fish

The disclosed transgenic fish are produced by introducing a transgene construct into cells of a fish, preferably embryonic cells, and most preferably in a single cell embryo. Where the transgene construct is introduced into embryonic cells, the transgenic fish is obtained by allowing the embryonic cell or cells to develop into a fish. Introduction of constructs into embryonic cells of fish, and subsequent development of the fish, are simplified by the fact that embryos develop outside of the parent fish in most fish species.

The disclosed transgene constructs can be introduced into embryonic fish cells using any suitable technique. Many techniques for such introduction of exogenous genetic material have been demonstrated in fish and other animals. These include microinjection (described by, for example, Culp *et al.* (1991)), electroporation (described by, for example, Inoue *et al.*, *Cell. Differ. Develop.* 29:123-128 (1990); Müller *et al.*, *FEBS Lett.* 324:27-32 (1993); Murakami *et al.*, *J. Biotechnol.* 34:35-42 (1994); Müller *et al.*, *Mol. Mar. Biol. Biotechnol.* 1:276-281 (1992); and Symonds *et al.*, *Aquaculture* 119:313-327 (1994)), particle gun bombardment (Zelenin *et al.*, *FEBS Lett.* 287:118-120 (1991)), and the use of liposomes (Szelei *et al.*, *Transgenic Res.* 3:116-119 (1994)). Microinjection is preferred. The preferred method for introduction of transgene constructs into fish embryonic cells by microinjection is described in the examples.

Embryos or embryonic cells can generally be obtained by collecting eggs immediately after they are laid. Depending on the type of fish, it is generally preferred that the eggs be fertilized prior to or at the time of collection. This is preferably accomplished by placing a male and
5 female fish together in a tank that allows egg collection under conditions that stimulate mating. After collecting eggs, it is preferred that the embryo be exposed for introduction of genetic material by removing the chorion. This can be done manually or, preferably, by using a protease such as pronase. A preferred technique for collecting zebrafish eggs and
10 preparing them for microinjection is described in the examples. A fertilized egg cell prior to the first cell division is considered a one cell embryo, and the fertilized egg cell is thus considered an embryonic cell.

After introduction of the transgene construct the embryo is allowed to develop into a fish. This generally need involve no more than
15 incubating the embryos under the same conditions used for incubation of eggs. However, the embryonic cells can also be incubated briefly in an isotonic buffer. If appropriate, expression of an introduced transgene construct can be observed during development of the embryo.

Fish harboring a transgene can be identified by any suitable
20 means. For example, the genome of potential transgenic fish can be probed for the presence of construct sequences. To identify transgenic fish actually expressing the transgene, the presence of an expression product can be assayed. Several techniques for such identification are known and used for transgenic animals and most can be applied to
25 transgenic fish. Probing of potential or actual transgenic fish for nucleic acid sequences present in or characteristic of a transgene construct is preferably accomplished by Southern or Northern blotting. Also preferred is detection using polymerase chain reaction (PCR) or other sequence-specific nucleic acid amplification techniques. Preferred
30 techniques for identifying transgenic zebrafish are described in the examples.

4. Identifying the Pattern of Expression of Fish Genes

Identifying the pattern of expression in the disclosed transgenic fish can be accomplished by measuring or identifying expression of the transgene in different tissues (tissue-specific expression), at different times
5 during development (developmentally regulated expression or developmental stage-specific expression), in different cell lineages (cell lineage-specific expression). These assessments can also be combined by, for example, measuring expression (and observing changes, if any) in a cell lineage during development. The nature of the expression product to
10 be detected can have an effect on the suitability of some of these analyses. On one level, different tissues of a fish can be dissected and expression can be assayed in the separate tissue samples. Such an assessment can be performed when using almost any expression product. This technique is commonly used in transgenic animals and is useful for assessing tissue-
15 specific expression.

This technique can also be used to assess expression during the course of development by assaying for the expression product at different developmental stages. Where detection of the expression product requires fixing of the sample or other treatments that destroy or kill the developing
20 embryo or fish, multiple embryos must be used. This is only practical where the expression pattern in different embryos is expected to be the same or similar. This will be the case when using the disclosed transgenic fish having stable and predictable expression.

A more preferred way of assessing the pattern of expression of a transgene during development is to use an expression product that can be
25 detected in living embryos and animals. A preferred expression product for this purpose is the green fluorescent protein. A preferred form of GFP and a preferred technique for measuring the presence of GFP in living fish is described in the examples.

30 Expression products of the disclosed transgene constructs can be detected using any appropriate method. Many means of detecting expression products are known and can be applied to the detection of

expression products in transgenic fish. For example, RNA can be detected using any of numerous nucleic acid detection techniques. Some of these detection methods as applied to transgenic fish are described in the examples. The use of reporter proteins as the expression product is
5 preferred since such proteins are selected based on their detectability. The detection of several useful reporter proteins is described by Iyengar *et al.* (1996).

In zebrafish, the nervous system and other organ rudiments appear within 24 hours of fertilization. Since the nearly transparent
10 zebrafish embryo develops outside its mother, the origin and migration of lineage progenitor cells can be monitored by following expression of an expression product in transgenic fish. In addition, the regulation of a specific gene can be studied in these fish.

Using zebrafish promoters that drive expression in specific
15 tissues, a number of transgenic zebrafish lines can be generated that express a reporter protein in each of the major tissues including the notochord, the nervous system, the brain, the thymus, and in other tissues (see Table 1). Other important lineages for which specific expression can be obtained include neural crest, germ cells, liver, gut, and kidney.
20 Additional tissue specific transgenic fish can be generated by using "enhancer trap" constructs to identify expression sequences in fish.

Table 1

	Source of Expression Sequences	Tissues/Cell lineages
	GATA-1	Erythroid progenitor
5	GATA-2	Hematopoietic stem cells/CNS
	Tinman	Heart
	Rag-1	T and B Cells
	Globin	Mature red blood cells
	MEF	Muscle progenitors
10	Goosecoid	Dorsal organizer
	SCL-1	Hematopoietic stem cells
	Rbtn-2	Hematopoietic stem cells
	No-tail	Notochord
	Flk-1	Vascular endothelia
15	Eve-1	Ventral/posterior cells
	Ikaros	Early lymphoid progenitors
	Pdx-1	Pancreas
	Islet-1	Motoneuron
	Shh	Multi-tissue induction/Left-right symmetry
20	Twist	Axial mesoderm/Left-right symmetry
	Krox20	Brain
	BMP4	Ventral mesoderm induction

5. Identifying Compounds That Affect Expression of Fish Genes

For many genes, and especially for genes involved in
 25 developmental processes, it would be useful to identify compounds that
 affect expression of the genes. The disclosed transgenic fish can be
 exposed to compounds to assess the effect of the compound on the
 expression of a gene of interest. For example, test compounds can be
 administered to transgenic fish harboring an exogenous construct
 30 containing the expression sequences of a fish gene of interest operably
 linked to a sequence encoding a reporter protein. By comparing the
 expression of the reporter protein in fish exposed to a test compound to

those that are not exposed, the effect of the compound on the expression of the gene from which the expression sequences are derived can be assessed.

6. Identifying Genes That Affect Expression of Fish Genes

5 Numerous mutants have been generated and characterized in zebrafish which collectively affect most developmental processes. The disclosed transgenic fish can be used in combination with these and other mutations to assess the effect of a mutant gene on the expression of a gene of interest. For example, mutations can be introduced into strains of
10 transgenic fish harboring an exogenous construct containing the expression sequences of a fish gene of interest operably linked to a sequence encoding a reporter protein. By comparing the expression of the reporter protein in fish with a mutation to those without the mutation, the effect of the mutation on the expression of the gene from which the
15 expression sequences are derived can be assessed.

The effect of such mutations on specific developmental processes and on the growth and development of specific cell lineages can also be assessed using the disclosed transgenic fish expressing a reporter protein in specific cell lineages or at specific developmental stages.

20 7. Genetically Marking Mutant Fish Genes

The disclosed transgene constructs can be used to genetically mark mutant genes or chromosome regions. For example, in zebrafish, recent chemical mutagenesis screens have generated more than one thousand different mutants with defects in most developmental processes.
25 If fish carrying a mutation generated in these screens could be more easily identified, a lot of time and labor would be saved. One way to promote rapid identification of fish carrying mutations would be the establishment of balancer chromosomes that carry markers that can be easily identified in living fish. This technology has greatly facilitated the task of
30 identification and maintenance of mutant stocks in *Drosophila* (Ashburner, *Drosophila, A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989); Lindsey and Zimm, *The Genome of*

Drosophila melanogaster (Academic Press, San Diego, CA, 1995)). As used herein, genetically marking a gene or chromosome region refers to genetically linking a reporter gene to the gene or chromosome region. Genetic linkage between two genetic elements (such as genes) refers to the elements being in sufficiently close proximity on a chromosome that they do not segregate from each other at random in genetic crosses. The closer the genetic linkage, the more likely that the two elements will segregate together. For genetic marking, it is preferred that the transgene construct segregate with the gene or chromosomal region of interest more than 60% of the time, it is more preferred that the transgene construct segregate with the gene or chromosomal region of interest more than 70% of the time, it is still more preferred that the transgene construct segregate with the gene or chromosomal region of interest more than 80% of the time, it is still more preferred that the transgene construct segregate with the gene or chromosomal region of interest more than 90% of the time, and it is most preferred that the transgene construct segregate with the gene or chromosomal region of interest more than 95% of the time.

Example 1 shows that living transgenic fish carrying insertions of a transgene, in which the zebrafish GATA-1 promoter has been ligated to the green fluorescent protein (GFP) reporter gene, can be identified by simple observation of GFP expression in blood cells. As in *Drosophila*, zebrafish chromosomal recombination occurs at a significantly lower rate during spermatogenesis than it does during oogenesis. Therefore, a transgene insertion that maps near a chemically induced mutant gene can be crossed into the mutant chromosome through oogenesis and will then remain linked to the mutation in male fish through many generations. This procedure will allow the identification of progeny harboring the mutant gene by simple observation of GFP in blood cells.

In the case of zebrafish, 200 lines carrying the GATA-1/GFP transgene (or another reporter construct), randomly inserted throughout the zebrafish genome should result in an average of 8 insertions in each of the 25 zebrafish chromosomes. This is possible since expression from the

disclosed constructs is not limited by effects of the site of insertion and the site of integration is not limited. The insertion sites can be mapped and then crossed through oogenesis into zebrafish lines that carry a mutation that maps nearby. Once established, mutant strains that carry
5 balancer chromosomes can be maintained in male fish.

Although it is preferred that mutant genes be genetically marked, any gene of interest or any chromosome region can be marked, and the maintenance and inheritance of the gene can be monitored, in a similar manner. As used herein, an identified mutant gene is a mutant gene that
10 is known or that has been identified, in contrast to a mutant gene which may be present in an organism but which has not been recognized.

Genetically mapping of mutant genes or transgenes in fish can be performed using established techniques and the principles of genetic crosses. Generally, mapping involves determining the linkage
15 relationships between genetic elements by assessing whether, and to what extent two or more genetic elements tend to cosegregate in genetic crosses.

8. Identifying Fish That Have Inherited a Mutant Gene

Mutant fish in which the mutant gene is marked with an
20 exogenous construct expressing a reporter protein simplify the identification of progeny fish that carry the mutant gene. For example, after a cross, progeny fish can be screened for expression of the reporter protein. Those that express the reporter protein are very likely to have inherited the mutant gene which is genetically linked. Those progeny fish
25 not expressing the reporter protein can be excluded from further analysis.

Although recombination during gametogenesis may result in segregation of the exogenous construct from the mutant gene, this will happen only rarely. Initial screening for fish expressing the reporter protein will still ensure that the majority of such progeny fish will carry
30 the mutant gene. Confirmation of the mutant can be established by subsequent direct testing for the mutant gene.

9. Identifying and Cloning Regulatory Sequences from Fish

The disclosed constructs can also be used as "enhancer traps" to generate transgenic fish that exhibit tissue-specific expression of an expression product. Transgenic animals carrying enhancer trap constructs
5 often exhibit tissue-specific expression patterns due to the effects of endogenous enhancer elements that lie near the position of integration.

Once it is determined that the exogenous construct is operably linked to an enhancer or other regulatory sequence in a fish, the regulatory element can be isolated by re-cloning the transgene construct.
10 Many general cloning techniques can be used for this purpose. A preferred method of cloning regulatory sequences that have become linked to a transgene construct in a fish is to isolate and cleave genomic DNA from the fish with a restriction enzyme that does not cleave the exogenous construct. The resulting fragments can be cloned *in vitro* and screened
15 for the presence of characteristic transgene sequences. A search for enhancers in zebrafish using a transgene construct having only a promoter operably linked to a sequence encoding a reporter protein has generated a transgenic line that expresses GFP exclusively in hatching gland cells.

A similar procedure can be followed to identify promoters. In
20 this case, a "promoter probe" construct, which lacks any expression sequences, is used. Only if the construct is inserted into the genome downstream of expression sequences will the expression product encoded by the construct be expressed.

10. Identifying Promoters and Enhancers in Cloned Expression 25 Sequences

The linked genomic sequences of clones identified as containing expression sequences, or any other nucleic acid segment containing expression sequences, can then be characterized to identify potential and actual regulatory sequences. For example, a deletion series of a positive
30 clone can be tested for expression in transgenic fish. Sequences essential for expression, or for a pattern of expression, are identified as those which, when deleted from a construct, no longer support expression or

the pattern of expression. The ability to assess the pattern of expression of a transgene in fish using the disclosed transgenic fish and methods makes it possible to identify the elements in the regulatory sequences of a fish gene that are responsible for the pattern of expression. The disclosed
5 transgenic fish, since they can be produced routinely and consistently, allow meaningful comparison of the expression of different deletion constructs in separate fish.

An example of the power of this capability is described in Example 2. Application of this system to the study of the GATA-2
10 promoter has led to identification of enhancer regions that facilitate gene expression specifically in hematopoietic precursors, the enveloping layer (EVL) and the central nervous system (CNS). Through site-directed mutagenesis, it has been discovered that the DNA sequence CCCTCCT is essential for the neuron-specific activity of the GATA-2 promoter. This
15 is described in Example 2.

11. Isolating Cells Expressing An Expression Product

Using cell sorting based on the presence of an expression product, pure populations of cells expressing a transgene construct can be isolated from other cells. Where the transgene construct is expressed in particular
20 cell lineages or tissues, this can allow the purification of cells from that particular lineage. These cells can be used in a variety of *in vitro* studies. For instance, these pure cell populations can provide mRNA for differential display or subtractive screens for identifying genes expressed in that cell lineage. Progenitor cells of specific tissue could also be
25 isolated. Establishing such cells in tissue culture would allow the growth factor needs of these cells to be determined. Such knowledge could be used to culture non-transgenic forms of the same cells or related cells in other organisms.

Cell sorting is preferably facilitated by using a construct
30 expressing a fluorescent protein or an enzyme producing a fluorescent product. This allows fluorescence activated cell sorting (FACS). A preferred fluorescent protein for this purpose is the green fluorescent

protein. The ability to generate transgenic fish expressing GFP in a tissue- and cell lineage-specific manner for different cell types indicates that transgenic fish that express GFP in other types of tissues can be generated in a straightforward manner. The disclosed FACS approach
5 can therefore be used as a general method for isolating pure cell populations from developing embryos based solely on gene expression patterns. This method for isolation of specific cell lineages is preferably performed using constructs linking GFP with the expression sequences of genes identified as being involved in development. Numerous such genes
10 have been or can be identified as mutants that affect development. Cells isolated in this manner should be useful in transplantation experiments.

Examples

15 **Example 1: Tissue-specific Expression and Germline Transmission of a Transgene in Zebrafish.**

In this example, DNA constructs containing the putative zebrafish expression sequences of GATA-1, an erythroid-specific transcription factor, operatively linked to a sequence encoding the green fluorescent protein (GFP), were microinjected into single-cell zebrafish embryos.

20 GATA-1, an early marker of the erythroid lineage, was initially identified through its effects upon globin gene expression (Evans and Felsenfeld, *Cell* 58:877-85 (1989); Tsai *et al.*, *Nature* 339:446-51 (1989)). Since then GATA-1 has been shown to be a member of a multigene family. Members of this gene family encode transcription
25 factors that recognize the DNA core consensus sequence, WGATAR (SEQ ID NO:18). GATA factors are key regulators of many important developmental processes in vertebrates, particularly hematopoiesis (Orkin, *Blood* 80:575-81 (1992)). The importance of GATA-1 for hematopoiesis was definitively demonstrated in null mutations in mouse (Pevny *et al.*,
30 *Nature* 349:257-60 (1991)). In chimeric mice, embryonic stem cells carrying a null mutation in GATA-1, created via homologous

recombination, contributed to all non-hematopoietic tissues tested and to a white blood cell fraction, but failed to give rise to mature red blood cells.

In zebrafish, GATA-1 expression is restricted to erythroid progenitor cells that initially occupy a ventral extra-embryonic position,
5 similar to the situation found in other vertebrates (Detrich *et al.*, *Proc Natl Acad Sci U S A* **92**:10713-7 (1995)). As development proceeds, these cells enter the zebrafish embryo and form a distinct structure known as the hematopoietic intermediate cell mass (ICM).

Vertebrate hematopoiesis is a complex process that proceeds in
10 distinct phases, at various anatomic sites, during development (Zon, *Blood* **86**:2876-91 (1995)). Although studies on *in vitro* model systems have generated some insight into hematopoietic development (Cumano *et al.*, *Cell* **86**:907-16 (1996); Kennedy *et al.*, *Nature* **386**:488-493 (1997); Medvinsky and Dzierzak, *Cell* **86**:897-906 (1996); Nakano *et al.*, *Science*
15 **272**:722-4 (1996)), the origin of hematopoietic progenitor cells during vertebrate embryogenesis is still controversial. Therefore, an *in vivo* model should be useful to determine precisely the cellular and molecular mechanisms involved in hematopoietic development. Such a model could also be used to identify compounds and genes that affect hematopoiesis.
20 In mammals, since embryogenesis occurs internally, it is difficult to carefully observe hematopoietic processes.

Zebrafish have a number of features that facilitate the study of vertebrate hematopoiesis. Because development is external and embryos are nearly transparent, the migration of labeled hematopoietic cells can be
25 easily monitored. In addition, many mutants that are defective in hematopoietic development have been generated (Ransom *et al.*, *Development* **123**:311-319 (1996); Weinstein *et al.*, *Development* **123**:303-309 (1996)). Zebrafish embryos that significantly lack circulating blood can survive for several days, so downstream effects of mutations upon
30 gene expression deleterious to embryonic hematopoietic development can be characterized. Since the cellular processes and molecular regulation of hematopoiesis are generally conserved throughout vertebrate evolution,

results from zebrafish embryonic studies can also provide insight into the mechanisms involved in mammalian hematopoiesis.

Cloning and sequencing of GATA-1 genomic DNA

A zebrafish genomic phage library was screened with a ³²P
5 radiolabeled probe containing a region of zebrafish GATA-2 cDNA that encodes a conserved zinc finger. A number of positive clones were identified. The inserts in these clones were cut with various restriction enzymes. The resulting fragments were subcloned into pBluescript II
10 KS(-) and sequenced. Based on DNA sequence analysis, two phage clones were shown to contain zebrafish GATA-1 sequences. The cDNA sequence of zebrafish GATA-1 is described by Detrich *et al.*, *Proc. Natl. Acad. Sci. USA* 92:10713 (1995). Nucleotide sequence of the GATA-1 promoter region is shown in SEQ ID NO:26.

Plasmid constructs

15 Construct G1-(Bgl)-GM2 was generated by ligating a modified GFP reporter gene (GM2) to a 5.4 kb *EcoRI/BglIII* fragment that contains putative zebrafish GATA-1 expression sequences, that is, the 5' flanking sequences upstream of the major GATA-1 transcription start site. GM2 contains 5' wild type GFP and a 3' *NcoI/EcoRI* fragment derived from a
20 GFP variant, m2, that emits approximately 30 fold greater fluorescence than does the wild type GFP under standard FITC conditions (Cormack *et al.*, *Gene* 173:33-8 (1996)). This construct is illustrated as construct (1) in Figure 2.

To isolate expression sequences in the 5' untranslated region of
25 GATA-1, a 5.6 kb DNA fragment was amplified by the polymerase chain reaction (PCR) from a GATA-1 genomic subclone using a T7 primer which is complementary to the vector sequence, and a specific primer, Oligo (1), that is complementary to the cDNA sequence just 5' of the GATA-1 translation start. The GATA-1 specific primer contained a
30 *BamHI* site to facilitate subsequent cloning. The PCR reaction was performed using Expand™ Long Template PCR System (Boehringer Mannheim) for 30 cycles (94°C, 30 seconds; 60°C, 30 seconds; 68°C, 5

minutes). After digestion with *BamHI* and *XhoI*, this 5.6 kb DNA fragment was gel purified and ligated to DNA encoding the modified GFP, resulting in construct G1-GM2 (construct (2) in Figure 2). The construct G1-(5/3)-GM2 was generated by ligating an additional 4 kb of
5 GATA-1 genomic sequences, which contains GATA-1 intron and exon sequences, to the 3' end (following the polyadenylation signal) of the reporter gene in construct G1-GM2. This construct is illustrated as construct (3) in Figure 2.

Fish and Microinjection

10 Wild type zebrafish embryos were used for all microinjections. The zebrafish were originally obtained from pet shops (Culp *et al.*, *Proc Natl Acad Sci U S A* 88:7953-7 (1991)). Fish were maintained on reverse osmosis-purified water to which Instant Ocean (Aquarium Systems, Mentor, OH.) was added (50 mg/l). Plasmid DNA G1-GM2 was
15 linearized using restriction enzyme *AatII* (which cuts in the vector backbone), while plasmid DNA G1-(5/3)-GM2 was excised from the vector by digestion with restriction enzyme *SacI*, and separated using a low melting agarose gel. DNA fragments were cleaned using GENECLAN II Kit (Bio101 Inc.) and resuspended in 5 mM Tris, 0.5
20 mM EDTA, 0.1 M KCl at a final concentration of 50 μ g/ml prior to microinjection. Single cell embryos were prepared and injected as described by Culp *et al.*, *Proc Natl Acad Sci U S A* 88:7953-7 (1991), except that tetramethyl-rhodamine dextran was included as an injection control. This involved collecting newly fertilized eggs, dechorionating
25 the eggs with pronase (used at 0.5 mg/ml), and injecting DNA. Injection with each construct was done independently 5 to 10 times and the data obtained were pooled.

Fluorescent microscopic observation and imaging

Embryos and adult fish were anesthetized using tricaine (Sigma
30 A-5040) as described previously (Westerfield, *The Zebrafish Book* (University of Oregon Press, 1995)) and examined under a FITC filter on a Zeiss microscope equipped with a video camera. Images of circulating

blood cells were produced by printing out individual frames of recorded videos. Other pictures of fluorescent embryos were generated by superimposing a bright field image on a fluorescent image using Adobe Photoshop software. One month old fish were anesthetized and then
5 rapidly embedded in OCT. Sections of 60 μ m were cut using a cryostat and were immediately observed by fluorescence microscopy.

Identification of germline transgenic fish by PCR

DNA isolation, internal control primers and PCR conditions were the same as described by Lin *et al.* *Dev Biol* 161:77-83 (1994)). Briefly,
10 DNA was extracted from pools of 40 to several hundred dechorionated embryos (obtained from mating a single pair of fish) at 16 to 24 hours of development by vortexing for 1 minute in a buffer containing 4 M guanidium isothiocyanate, 0.25 mM sodium citrate (pH 7.0), and 0.5% Sarkosyl, 0.1 M β -mercaptoethanol. The sample was extracted once with
15 phenol:chloroform: isoamyl alcohol (25:24:1) and total nucleic acid was precipitated by the addition of 3 volumes of ethanol and 1/10 volume sodium acetate (3 M, pH 5.5). The pellet was washed once in 70% ethanol and dissolved in 1X TE (pH 8.0).

Approximately 0.5 μ g of DNA was used in a PCR reaction
20 containing 20 mM Tris (pH 8.3), 1.5 mM $MgCl_2$, 25 mM KCl, 100 μ g/ml gelatin, 20 pmole each PCR primer, 50 μ M each dNTPs, 2.5 U Taq DNA polymerase (Pharmacia). The reaction was carried out at 94°C for 2.5 minutes for 30 cycles with a 5 minute initial 94°C denaturation step, and a 7 minute final 72°C elongation step. Specific primers, Oligos
25 (2) and (3), that were used to detect GFP, generated a 267 bp product. A pair of internal control primers homologous to sequences of the zebrafish homeobox gene, ZF-21 (Njolstad *et al.*, *FEBS Letters* 230:25-30 (1988)), was included in each reaction. This pair of primers should generate a PCR product of 475 bp for all PCR reactions using zebrafish DNA.

30 Preparation of embryonic cells and flow cytometry

Embryos were disrupted in Holfereter's solution using a 1.5 ml pellet pestle (Kontes Glass, OEM749521-1590). Cells were collected by

centrifugation (400 g, 5 minutes). After digestion with 1X Trypsin/EDTA for 15 minutes at 32°C, the cells were washed twice with phosphate buffered saline (PBS) and filtered through a 40 micron nylon mesh. Fluorescence activated cell sorting (FACS) was performed under standard FITC conditions.

cDNA synthesis and PCR

Total RNA was extracted from FACS purified cells using the RNA isolation kit, TRIzol (Bio101). Reverse transcription and PCR (RT-PCR) were performed using the Access RT-PCR System from Promega (Catalog # A1250). Specific primers, Oligos (4) and (5), used to detect the zebrafish GATA-1 cDNA, generated a 410 bp product.

Oligonucleotides

- (1) 5'-CCGGATCCTGCAAGTGTAGTATTGAA-3' (GATA-1, promoter antisense; SEQ ID NO:1);
- (2) 5'-AATGTATCAATCATGGCAGAC-3' (GM2 sense; SEQ ID NO:2);
- (3) 5'-TGTATAGTTCATCCATGCCATGTG-3' (GM2 antisense; SEQ ID NO:3);
- (4) 5'-ATGAACCTTTCTACTCAAGCT-3' (GATA-1, cDNA sense; SEQ ID NO:4)
- (5) 5'-GCTGCTTCCACTTCCACTCAT-3' (GATA-1, cDNA antisense; SEQ ID NO:5)

Whole-mount RNA *in situ* hybridization

Sense and antisense digoxigenin-labeled RNA probes were generated from a GATA-1 genomic subclone containing the second and third exon coding sequence using a DIG/Genius™ 4 RNA Labeling Kit (SP6/T7) (Boehringer Mannheim). RNA *in situ* hybridizations were performed as described (Westerfield, The Zebrafish Book (University of Oregon Press, 1995)).

Genomic structure of the zebrafish GATA-1

Two clones containing zebrafish GATA-1 sequences were isolated from a lambda phage zebrafish genomic library as described above.

Restriction enzyme mapping indicated that the two overlapping clones contained approximately 35 kb of the GATA-1 locus. To define the promoter of the zebrafish GATA-1 gene, transcription initiation sites for the zebrafish GATA-1 were mapped by primer extension. As in chicken,
5 mouse, human and other species, multiple transcription initiation sites were identified. A major transcription initiation site was mapped 187 bases upstream of the translation start.

Comparison of the GATA-1 genomic structure for human, mouse and chicken suggested that the intron-exon junction sequences of this gene
10 are likely to be conserved throughout vertebrates. Oligonucleotide primers flanking potential GATA-1 introns were designed and used to sequence the zebrafish genomic clones. Sequence analysis revealed that the zebrafish GATA-1 gene consists of five exons and four introns which lie within a 6.5 kb genomic region (Figure 1). Although the exon-intron
15 number and junction sequences are well conserved between zebrafish and other vertebrates, the zebrafish GATA-1 introns are smaller than in other species.

Transient expression of GFP driven by the GATA-1 promoter in zebrafish embryos

20 Based on the zebrafish GATA-1 genomic structure, three GFP reporter gene constructs were generated (Figure 2). Construct G1-(Bgl)-GM2 was generated by ligation of a modified GFP reporter gene (GM2) to a 5.4 kb *EcoRI/BglIII* fragment that contains the 5' flanking sequences upstream of the major GATA-1 transcription start site.
25 Construct G1-GM2 contained a 5.6 kb region upstream of the translation start of GATA-1. The third construct, G1-(5/3)-GM2, was generated by ligating an additional 4 kb of GATA-1 genomic sequences, which contain intron and exon sequences, to the 3' end of the reporter gene in construct G1-GM2. Each construct was microinjected into the cytoplasm of single
30 cell zebrafish embryos. GFP reporter gene expression in the embryos was examined at a number of distinct developmental stages by fluorescence microscopy.

GFP expression was observed in embryos injected with either construct G1-GM2 or construct G1-(5/3)-GM2 as early as 80% epiboly, approximately 8 hours post fertilization (pf). At that time, GFP positive cells were restricted to the ventral region of the injected embryos. At 16
5 hours pf, GFP expression was clearly visible in the developing intermediate cell mass (ICM), the earliest hematopoietic tissue in zebrafish. After 24 hours pf, GFP positive cells were observed in circulating blood and could be continuously observed in circulating blood for several months. During the first five days pf, examination of
10 circulating blood revealed two distinct cell populations with different levels of GFP expression. One cell type was larger and brighter; the other smaller and less bright. No significant difference in GFP expression levels was detected between embryos injected with either construct G1-GM2 or G1-(5/3)-GM2. However, injection of construct
15 G1-(Bgl)-GM2 yielded very weak GFP expression in developing embryos. This result indicated that either the GATA-1 transcription initiation site was removed by *Bgl*III restriction digestion, or that the 5' untranslated region of zebrafish GATA-1 is required for high level tissue specific expression of GFP. It is not surprising that a construct lacking the 5'
20 untranslated region of GATA-1 did not generate much GFP expression in microinjected embryos. These regions are often needed for transcript stability. At times, these regions also contain binding sites for regulators of gene expression.

At least 75% of the embryos injected with G1-GM2 or
25 G1-(5/3)-GM2 construct showed some degree of ICM specific GFP expression (Table 2). The number of GFP positive cells in the ICM or in circulation ranged from a single cell to a few hundred cells. Less than 7% of these embryos showed GFP expression in non-hematopoietic tissues, usually limited to fewer than ten cells per embryo. Non-specific
30 expression of GFP was usually observed in the notochord, muscle, and enveloping cell layers, and was limited to no more than 10 cells per embryo. These observations indicated that a genomic GATA-1 fragment

extending approximately 5.6 kb upstream from the GATA-1 translation start site ligated to GFP sufficed to recapitulate the embryonic pattern of GATA-1 expression in zebrafish.

Table 2

5	Constructs	No. observed embryos	No. embryos with GFP expression in ICM (%)	No. embryos with strong GFP expression in ICM (%) ^a	No. embryos with non- specific expression GFP (%)
	G1-GM2	336	274 (81.5%)	177 (52.7%)	15 (4.5%)
	G1-GM2(5/3)	248	187 (75.4%)	150 (60.5%)	16 (6.5%)
	G1(Bg1II)-GM2	370	0 (0%)	0 (0%)	19 (5.1%)

^aStrong GFP expression means that each embryo has more than 10 green fluorescent cells in the ICM.

GFP expression in germline GATA-1/GFP transgenic zebrafish

Microinjected zebrafish embryos were raised to sexual maturity and mated. Progeny were tested by PCR to determine the frequency of germline transmission of the GATA-1/GFP transgene. Nine of six
 15 hundred and seventy two founder fish have transmitted GFP to the F1 generation. Examination of these fish by fluorescence microscopy revealed that seven of eight lines expressed GFP in the ICM and in circulating blood cells. GFP expression patterns in the ICM were consistent with the RNA *in situ* hybridization patterns previously observed
 20 for GATA-1 mRNA expression in zebrafish (Detrich *et al.*, *Proc Natl Acad Sci U S A* 92:10713-7 (1995)). In the two lines where F2 transgenic fish have been obtained, GFP expression in blood cells was observed in 50% of the progeny when a transgenic F2 was mated to a non-transgenic fish. This indicated that GFP was transmitted to progeny in a Mendelian
 25 fashion. Southern blot analysis showed that GFP transgene insertions

occurred at different sites in these two lines. In one line, transgenic fish apparently carry 4 copies of the transgene and in the other line, 7 copies.

Blood cells were collected from 48 hour transgenic fish by heart puncture and a blood smear was observed by fluorescence microscopy.

- 5 Two distinct populations of fluorescent cells were observed in these smears. As in the circulation of embryos that transiently express GFP, one cell population was observed that was large and bright and another that was smaller and less bright. Although the blood cells collected from adult transgenic zebrafish showed some variability in fluorescence
10 intensity, they appeared to have uniform size. Blood cells collected from non-transgenic fish showed no fluorescence.

In two day old transgenic zebrafish, weak GFP expression was observed in the heart. GFP expression was also observed in the eyes and, in three of seven transgenic lines, in some neurons of the spinal cord.

- 15 Expression in the eyes peaked between 30 and 48 hours pf and became extremely weak by day 4. It is thought that expression of GFP in eyes and neurons may replicate the authentic GATA-1 expression pattern.

- Examination of GFP expression in tissues of one month old fish showed that the head kidney contained a large number of fluorescent
20 cells. This result suggests that the kidney is the site of adult erythropoiesis in zebrafish. It has been reported that GATA-1 is expressed in the testes of mice. Expression of GFP was not found in testes dissected from adult fish. It is possible that the disclosed GATA-1 transgene constructs lack an enhancer required for testis expression of
25 GATA-1. Other tissues including brain, muscle and liver had no detectable level of GFP expression.

FACS analysis of GATA-1/GFP transgenic fish

- GFP expression in GATA-1/GFP transgenic fish allowed isolation of a pure population of the earliest erythroid progenitor cells for *in vitro*
30 studies by fluorescence activated cell sorting. F1 transgenic embryos were collected at the onset of GFP expression and cell suspensions were prepared. Approximately 3.6% of the cell populations of whole

transgenic fish were fluorescence positives as compared to 0.12% in the non-transgenic controls. Based on the number of embryos used, FACS analysis suggested that there are approximately three hundred erythroid progenitor cells per embryo at 14 hours pf.

5 To determine whether the FACS purified cells are enriched for GATA-1, RNA was isolated from these cells and GATA-1 mRNA levels were determined by RT-PCR. The results indicated that these cells were highly enriched for GATA-1 mRNA.

10 Erythroid specific expression was observed in living embryos during early development. Fluorescent circulating blood cells were detected in microinjected embryos 24 hours after fertilization and could still be observed in two month old fish. Germline transgenic fish obtained from the injected founders continued to express GFP in erythroid cells in the F1 and F2 generations. The GFP expression patterns in
15 transgenic fish were consistent with the RNA *in situ* hybridization pattern generated for GATA-1 mRNA expression. These transgenic fish allowed isolation, by fluorescence activated cell sorting, the earliest erythroid progenitor cells from developing embryos. Using constructs containing other zebrafish promoters and GFP, it will be possible to generate
20 transgenic fish that allow continuous visualization of the origin and migration of any lineage specific progenitor cells in a living embryo.

 The results described in this example indicate that monitoring GFP expression can be a more sensitive method than RNA *in situ* detection by which to determine gene expression patterns. For instance,
25 in the disclosed GATA-1/GFP transgenic fish, GFP expression in circulating blood allowed two types of cells to be distinguished. One cell type was larger and brighter; the other smaller and less bright. There were fewer of the larger, brighter cell type. These cells are believed to be erythroid precursors while the more abundant, smaller cells are
30 believed to be fully differentiated erythrocytes. Preliminary cell transplantation experiments with embryonic blood cells have shown that they contain a cell population that has long-term proliferation capacity.

In two day old transgenic zebrafish, GFP expression was observed in the heart. In adult transgenic zebrafish, GFP expression was observed in the kidney. By histological methods, it has been shown that the heart endocardium is a transitional site for hematopoiesis in embryonic zebrafish and that the kidney is the site of adult hematopoiesis (Al-Adhami and Kunz, *Develop. Growth and Differ.* 19:171-179 (1977)). The results in GATA-1/GFP transgenic fish support these observations.

The GFP expression seen in the eyes and neurons of embryonic transgenic fish may be due to a lack of a transcriptional silencer in the transgene constructs. It seems unlikely that the GFP expression in the eyes is due to positional effects caused by the sites of insertion since all seven transgenic lines have GFP expression in embryonic fish eyes.

Using fluorescence activated cell sorting, pure populations of hematopoietic progenitor cells were isolated from the ICM of transgenic zebrafish. Since approximately 10^7 cells can be sorted per hour, 10^5 to 10^6 purified ICM cells can be obtained in a few hours. These cells, which are derived from the earliest site of hematopoiesis in zebrafish, can be used in a variety of *in vitro* studies. For instance, these pure cell populations can provide mRNA for differential display or subtractive screens for identifying novel hematopoietic genes. Erythroid precursors obtained from the ICM might also be established in tissue culture. This would allow the growth factor needs of these cells to be determined.

The approach to obtaining and studying transgene expression in erythroid cells described above is generally applicable to the study of any developmentally regulated process. This approach can also be applied to the identification of cis-acting promoter elements that are required for tissue specific gene expression (see Example 2). The analysis of promoter activity in a whole animal is desirable since dynamic temporal and spatial changes in a cellular microenvironment can be only poorly mimicked *in vitro*. The ease of generating and maintaining a large number of transgenic zebrafish lines makes obtaining statistically significant results practical. Finally, transgenic zebrafish that express

GFP in specific tissues provide useful markers for identifying mutations that affect these lines in genetic screens. Given the genetic resources and embryological methods available for zebrafish, transgenic zebrafish exhibiting tissue-specific GFP expression is a very valuable tool for
5 dissecting developmental processes.

Example 2: Identification of Enhancers in GATA-2 Expression Sequences.

A large number of studies have shown that neuronal cell
10 determination in invertebrates occurs in progressive waves that are regulated by sequential cascades of transcription factors. Much less is known about this process in vertebrates. It was realized that an integrated approach combining embryological, genetic and molecular methods, such as that used to study neurogenesis in *Drosophila* (Ghysen *et al.*, *Genes &*
15 *Dev* 7:723-33 (1993)), would facilitate the identification of the molecular mechanisms involved in specifying neuronal fates in vertebrates. The following is an example of identification of cis-acting sequences that control neuron-specific gene expression in a vertebrate. Such identification is an initial step toward unraveling similar cascades in a
20 vertebrate.

Transcription factors bind to cis-acting DNA sequences (sometimes referred to as response sequences) to regulate transcription. Often these transcription factors are members of multigene families that have overlapping, but distinct, expression patterns and functions. The
25 transcription factor GATA-2 is a member of such a gene family (Yamamoto *et al.*, *Genes Dev* 4:1650-62 (1990)). Each member of the GATA gene family is characterized by its ability to bind to cis-acting DNA elements with the consensus core sequence WGATAR (Orkin, *Blood* 80:575-81 (1992); SEQ ID NO:18). All protein products of the
30 GATA family contain two copies of a highly conserved structural motif, commonly known as a zinc finger, which is required for DNA binding (Martin and Orkin, *Genes Dev* 4:1886-98 (1994)). Six members of the

GATA family have been identified in vertebrates (Orkin, *Blood* **80**:575-81 (1992), Orkin, *Curr Opin Cell Biol* **7**:870-7 (1995)). *Pannier*, another member of the GATA gene family, is expressed in *Drosophila* neuronal precursors and inhibits expression of *achaete-scute*, a gene complex that
5 plays a critical role in neurogenesis in *Drosophila* (Romain *et al.*, *Development* **119**:1277-91 (1993)).

In chicken and mouse, the transcription factor GATA-2 is expressed in hematopoietic precursors, immature erythroid cells, proliferating mast cells, the central nervous system (CNS), and
10 sympathetic neurons (Yamamoto *et al.*, *Genes & Dev* **4**:1650-62 (1990), Orkin, *Blood* **80**:575-81 (1992), Jippo *et al.*, *Blood* **87**:993-8 (1996)). Studies in zebrafish (Detrich *et al.*, *Proc Natl Acad Sci U S A* **92**:10713-7 (1995)) and *Xenopus* (Zon *et al.*, *Proc Natl Acad Sci U S A* **88**:19642-6 (1991), Kelley *et al.*, *Dev Biol* **165**:193-205 (1994)) have also shown
15 that GATA-2 expression is restricted to hematopoietic tissues and the CNS. Homozygous null mutants, created in mouse via homologous recombination, have profound deficits in all hematopoietic lineages (Tsai *et al.*, *Nature* **371**:221-6 (1994)). The role played by GATA-2 in neuronal tissue of these mice has not been carefully examined, perhaps
20 because the embryos die before day E11.5. Analysis of GATA-2 expression in chick embryonic neuronal tissue after notochord ablation has suggested that GATA-2 plays a role in specifying a neurotransmitter phenotype (Groves *et al.*, *Development* **121**:887-901 (1995)). In addition, GATA factors are required for activity of the neuron-specific enhancer of
25 the gonadotropin-releasing hormone gene (Lawson *et al.*, *Mol Cell Biol* **16**:3596-605 (1996)).

The effects of various hematopoietic growth factors on GATA-2 expression has been carefully studied in tissue culture systems (Weiss *et al.*, *Exp Hematol* **23**:99-107 (1995)) and some growth factors have been
30 shown to have dramatic effects on early embryonic GATA-2 expression (Walmsley *et al.*, *Development* **120**:2519-29 (1994), Maeno *et al.*, *Blood* **88**:1965-72 (1996)). In addition, nuclear translocation of a maternally

supplied CCAAT binding transcription factor has been shown to be necessary for the onset of GATA-2 transcription at the mid-blastula transition in *Xenopus* (Brewer *et al.*, *Embo J* 14:757-66 (1995)).

However, prior to the disclosed work, nothing was known about the mechanisms that control neuron-specific expression of this gene.

Cloning and sequencing of 5' part of GATA-2 genomic DNA

A zebrafish genomic phage library was screened with the conserved zinc finger domain of zebrafish GATA-2 cDNA radiolabeled with ³²P. Two positive clones, λGATA-21 and λGATA-22, were identified. Restriction fragments of λGATA-21 were subcloned into pBluescript II KS(-). DNA sequence of the resulting clones was obtained from -4807 to +2605 relative to the GATA-2 translation start.

Nucleotide sequence of the GATA-2 promoter region is shown in SEQ ID NO:27. Unless otherwise indicated, positions within the GATA-2 clones use this numbering. The 7.3 kb region upstream of the translation start in λGATA-21 was amplified by the polymerase chain reaction (PCR) using Expand™ Long Template PCR System (Boehringer Mannheim) for 25 cycles (94°C, 30 seconds; 68°C, 8 minutes). Primers used were a T7 primer and a primer specific for sequences 5' to the GATA-2 translation start site (5'-ATGGATCCTCAAGTGTCCGCGCTTAGAA-3'; SEQ ID NO:19). The GATA-2 specific primer contained a *Bam*HI site to facilitate subsequent cloning. The PCR product (P1) was cloned into the *Sma*II/*Bam*HI sites of pBluescript II KS(-).

Plasmid constructs

The 7.3 kb DNA fragment containing the putative GATA-2 expression sequences (P1) was ligated to a modified GFP reporter gene (GM2, described above), resulting in construct P1-GM2 (Figure 3). Based on P1-GM2, constructs containing successive 5' deletions in the region upstream of the transcription start site were generated using the restriction sites *Pst*I, *Sac*I, *Aat*II, *Cla*I and *Sca*I in this upstream region (Figure 3). Constructs nsP5-GM2 and nsP6-GM2 were generated by ligating the 1116 bp fragment containing the GATA-2 neuron-specific

enhancer from -4807 to -3690 to P5-GM2 and P6-GM2, respectively (Figure 4). The same fragment containing the neuron-specific enhancer was also ligated to a 243 bp *SphI/BamHI* fragment of the *Xenopus* elongation factor 1 α (EF 1 α) minimal promoter that had previously been
 5 ligated to the GM2 gene, resulting in construct ns-XS-GM2 (Figure 4). The EF 1 α minimal promoter has been described in Johnson and Krieg, *Gene* 147:223-6 (1994).

PCR mapping of neuron-specific enhancer

PCR technology was exploited to create a deletion series within
 10 the 1116 bp neuron-specific enhancer using nsP5-GM2 as a template. A total of 10 specific 22-mer primers were synthesized. These included ns4647, ns4493, ns4292, ns4092, ns3990, ns3872, ns3851, ns3831, ns3800 and ns3789, in which the numbers refer to the positions of their 5' end base in the GATA-2 genomic sequence. A T7 primer was also used
 15 in the PCR reactions. The amplified fragments all contained the GM2 gene and SV40 polyadenylation signal in addition to the GATA-2 expression sequences. PCR reactions were performed using ExpandTM Long Template PCR System (Boehringer Mannheim) for 25 cycles (94°C, 30 seconds; 55°C, 30 seconds; 72°C, 2 minutes). The PCR products
 20 were purified with GENE CLEAN II Kit (Bio 101 Inc.) and subsequently used for microinjection.

After a 31 bp neural-specific enhancer was identified, five additional primers, each containing 2 or 3 mutant bases relative to the wild type enhancer sequence, were designed. These primers are (the
 25 mutant bases are underlined):

ns3831 5'-TCTGCGCCGCTTTCTGCCCCCTCCTGCCCTCTT-3' (SEQ ID NO:20)
 ns3831M1 5'-TCTGCGAAGCTTTCTGCCCCCTCCTGCCCTCTT-3' (SEQ ID NO:21)
 30 ns3831M2 5'-TCTGCGCCGCTTTCTGAACCCTCCTGCCCTCTT-3' (SEQ ID NO:22)
 ns3831M3 5'-TCTGCGCCGCTTTCTGCCAACTCCTGCCCTCTT-3' (SEQ ID NO:23)
 ns3831M4 5'-TCTGCGCCGCTTTCTGCCCCAAACTGCCCTCTT-3' (SEQ ID
 35 NO:24)

ns3831M5 5'-TCTGCGCCGCTTTCTGCCCCCTCCTGCCCTCTT-3' (SEQ ID NO:25)

These primers were used in conjunction with the T7 primer for PCR amplification of the target sequence using the nsP5-GM2 as the template.

- 5 PCR conditions were identical to those described above.

Microinjection of zebrafish

- Wild-type zebrafish were used for all microinjections. Plasmid DNA was linearized using single-cut restriction sites in the vector backbone, purified using GENECLAN II Kit (Bio 101 Inc.), and
10 resuspended in 5 mM Tris, 0.5 mM EDTA, 0.1 M KCl at a final concentration of 100 µg/ml. Single cell embryos were microinjected as described above. Each construct was injected independently 2 to 5 times and the data obtained were pooled.

Fluorescent microscopic observation

- 15 Embryos were anesthetized using tricaine as described above and examined under a FITC filter on a Zeiss microscope equipped with a video camera. Pictures showing GFP positive cells in living embryos were generated by superimposing a bright field image on a fluorescent image using Adobe Photoshop software.

Whole-mount RNA in situ hybridization

- 20 Sense and antisense digoxigenin-labeled RNA probes were generated from a GATA-2 cDNA subclone containing a 1 kb fragment of the 5' coding sequence using DIG/Genius™ 4 RNA Labeling Kit (SP6/T7) (Boehringer Mannheim). RNA *in situ* hybridizations were
25 performed as described by Westerfield (*The Zebrafish Book* (University of Oregon Press, 1995)).

Isolation of GATA-2 genomic DNA

- Two GATA-2 positive phage clones, λGATA-21 and λGATA-22, were identified as described above. Preliminary restriction analysis
30 suggested that λGATA-21 contained a large region upstream of the translation start codon. 7412 bp of this clone was sequenced from -4807 to +2605 relative to the translation start site. The putative GATA-2

expression sequences (P1) containing approximately 7.3 kb upstream of the translation start site from the λ GATA-21 was subcloned into a plasmid vector for expression studies.

Expression pattern of a modified GFP gene driven by the putative GATA-2 promoter in zebrafish embryos

The construct P1-GM2 was generated by ligation of a modified GFP reporter gene (GM2) to P1 (Figure 3). This construct was injected into the cytoplasm of single cell zebrafish embryos and GFP expression in the microinjected embryos was examined at a number of distinct developmental stages by fluorescence microscopy.

GFP expression was initially observed by fluorescence microscopy at the 4000 cell stage at about 4 hours post-injection (pi). At the dorsal shield stage (6 hours pi), GFP expression was observed throughout the prospective ventral mesoderm and ectoderm but expression in the dorsal shield was extremely rare. At 16 hours pi, GFP expression was observed in the developing intermediate cell mass (ICM), the early hematopoietic tissue of zebrafish. In addition, GFP expression could be seen in superficial EVL cells at 4 hours pi. Expression in the EVL peaked between 24 and 48 hours pi and became extremely weak by day 7. GFP expression in neurons, including extended axons, was first observed at 30 hours pi and was maintained at high levels through at least day 8.

Embryos injected with the P1-GM2 construct expressed GFP in a manner restricted to hematopoietic cells, EVL cells, and the CNS. The GFP expression patterns in gastrulating embryos, in the blood progenitor cells, and in neurons were consistent with the RNA *in situ* hybridization patterns previously generated for GATA-2 mRNA expression in zebrafish (Detrich *et al.*, *Proc Natl Acad Sci U S A* 92:10713-7 (1995)). However, GATA-2 expression in EVL has not been detected by RNA *in situ* hybridizations.

More than 95% of the embryos injected with P1-GM2 had tissue specific GFP expression (Table 3). About 5% of these embryos had non-specific GFP expression, limited to fewer than five cells per embryo.

These observations indicated that the DNA fragment extending approximately 7.3 kb upstream from the GATA-2 translation start site sufficed to correctly generate the embryonic tissue-specific pattern of GATA-2 gene expression.

5

Table 3

Construct	No. embryos observed	No. embryos with expression	No. embryos with circulating blood expression (%)	No. embryos with neuronal expression (%)	No. embryos with EVL expression (%)
P1-GM2	141	135	3 (2.13)	106 (75.2)	130 (92.2)
P2-GM2	198	177	32 (15.7)	136 (68.7)	175 (88.4)
P3-GM2	303	291	29 (9.6)	0 (0)	277 (91.4)
P4-GM2	143	126	21 (14.7)	0 (0)	118 (82.5)
P5-GM2	139	90	16 (11.5)	0 (0)	20 (14.4)
P6-GM2	138	44	2 (1.4)	0 (0)	11 (8.0)

Gross mapping of tissue-specific enhancers

To identify the portions of the GATA-2 expression sequences that are responsible for regulating tissue specific gene expression, several constructs containing deletions in the promoter were generated (Figure 3).

5 Naturally occurring restriction sites were used to create a series of gross deletions in the expression sequence region. Each construct was individually microinjected into single cell embryos. The developing embryos were observed by fluorescence microscopy at regular intervals for several days.

10 Embryos injected with P2-GM2, which contains GATA-2 sequences from -4807 to +1, expressed GFP in a manner similar to embryos injected with the original construct, P1-GM2 (Table 3). At 48 hr pi, GFP expression was observed in circulating blood cells, the CNS and the EVL. However, careful observation of the injected embryos at
 15 16 hr pi revealed that expression in the posterior end of the ICM was nearly abolished. This suggested that an enhancer for GATA-2 expression in early hematopoietic progenitor cells may reside in the

deleted region. Expression of GFP in circulating blood cells increased from approximately 2% to 16%, suggesting that a potential repressor for expression of GATA-2 in erythrocytes may also reside in the deleted region.

5 Embryos injected with P3-GM2, which contains GATA-2 sequences from -3691 to +1, expressed GFP in circulating blood cells and in the EVL, but did not express in the CNS. Embryos injected with other constructs that lack the deleted 1116 bp region, extending from -4807 to -3692, also had no GFP expression in the CNS (Table 3). It was
10 concluded that the 1116 bp region, extending from -4807 to -3692, contained a neuron-specific enhancer element.

 Embryos injected with P4-GM2, which contains GATA-2 sequences from -2468 to +1, had a GFP expression pattern similar to those injected with P3-GM2. Injection with P5-GM2, which contains
15 GATA-2 sequences from -1031 to +1, resulted in a sharp drop with respect to percentage of embryos expressing GFP in the EVL, but GFP expression in circulating blood cells was unaffected. This indicates that the 1437 bp region, extending from -2468 to -1032, contains an EVL-specific enhancer. The 1031 bp segment present in P5-GM2 may
20 represent the minimal expression sequences necessary for the maintenance of tissue specific expression of GATA-2.

Neuron-specific enhancer activity

 To confirm the neuron-specific enhancer activity of the 1116 bp region that spans from -4807 to -3692 of GATA-2, nsP5-GM2 was
25 constructed by ligating the 1116 bp fragment to P5-GM2, which contains the 1031 bp region upstream of the translation start of GATA-2 gene operably linked to a sequence encoding GM2 (Figure 4). Approximately 70% of the embryos injected with nsP5-GM2 had GFP expression in the CNS (Figure 5), while no embryos injected with P5-GM2 had GFP
30 expression in the CNS as noted in Table 3. This indicates that the 1116 bp region can effectively direct neuron-specific expression.

To determine whether the 1116 bp neuron-specific enhancer activity was context dependent, the construct ns-Xs-GM2 (Figure 4) was generated by ligating the enhancer to the *Xenopus* elongation factor 1 α minimal promoter (Johnson and Krieg, *Gene* 147:223-6 (1994)) operably
5 linked to the sequence encoding GM2 (Xs-GM2; Figure 4). When injected with Xs-GM2, embryos expressed GFP in various tissues including muscle, notochord, blood cells and melanocytes. However, no GFP expression was observed in the CNS (Figure 5). Injection with ns-XS-GM2 resulted in 8.5% of the embryos having GFP expression in the
10 CNS, far less than obtained by injection with nsP5-GM2 (Figure 5). Another construct, nsP6-GM2 (Figure 4), had an additional 653 bp deletion in the GATA-2 minimal expression sequence, extending from -1031 to -378. Injection of nsP6-GM2 resulted in 6.2% of embryos expressing GFP in the CNS (Figure 5). Injection with P6-GM2 resulted
15 in no GFP expression in the CNS (Table 3). These results suggests that the 1116 bp enhancer has some ability to confer neuronal specificity on a heterogeneous promoter, but requires proximal elements within its own promoter to exert its full activity.

20 Fine mapping of a neuron-specific cis-acting regulatory element

To precisely map the putative neuron-specific enhancer, a series of constructs containing progressive deletions in the 1116 bp DNA fragment was generated by PCR, using nsP5-GM2 as the template. The PCR products obtained were used directly for microinjection. The first
25 deletion series included ns4647, ns4493, ns4292, ns4092 and ns3990 (where the number indicates the upstream endpoint of the deleted fragment). Microinjection of all 5 mutants gave a similar percentage of embryos having GFP expression in the CNS (Figure 6). This indicated that a neuron-specific enhancer resides within the 298 bp sequence (from -
30 3990 to -3692) contained in ns3990.

Next, two additional deletion constructs, ns3872 and ns3789, were generated. As shown in Figure 6, over 60% of embryos injected with

ns3872 had GFP expression in the CNS, while embryos injected with ns3789 lacked GFP expression in the CNS. This indicated that the neuron-specific enhancer element was located within a 83 bp sequence from -3872 to -3790.

5 Injection of embryos with three additional deletion constructs ns3851, ns3831 and ns3800 allowed localization of the neuron-specific enhancer element to a 31 bp pyrimidine-rich sequence. This element has the sequence

5'-TCTGCGCCGCTTTCTGCCCCCTCCTGCCCTC-3' (nucleotides 1 to
10 31 of SEQ ID NO:20), which extends from -3831 to -3801 within the GATA-2 genomic DNA.

Site directed mutagenesis within neuron-specific enhancer element

To determine the core sequence necessary for the activity of the
15 neuron-specific element, five primers, each having two to three altered nucleotides within the 31 bp neuron-specific element (see above), were used to amplify nsP5-GM2. The PCR products obtained were directly injected into single cell embryos. This 31 bp sequence contains an Ets-like recognition site (AGGAC) in an inverted orientation which is present
20 in several neuron-specific promoters (Chang and Thompson, *J. Biol Chem* 271:6467-75 (1996), Charron *et al.*, *J. Biol Chem* 270:30604-10 (1995)). Therefore, four of the primers used in these PCR reactions contain altered nucleotides within the Ets-like recognition site or in the adjacent sequence. As expected, embryos injected with ns3831M1, which contains
25 two mutant nucleotides that are thirteen nucleotides upstream of the Ets-like recognition site, showed little change in neuron-specific GFP expression (Figure 7). A mutation of 2 nucleotides (ns3831M2) that lie three nucleotides upstream of the Ets-like recognition site had no effect on enhancer activity (Figure 7). Mutation of two nucleotides just one
30 nucleotide upstream of the Ets-like motif, contained in ns3831M3, completely eliminated the neuron-specific enhancer activity of the 31 bp element (Figure 7). Mutation of three nucleotides (ns3831M4), of which

two lie within the Ets-like recognition site, also resulted in a sharp decrease in enhancer activity (Figure 7). A mutation of two nucleotides that lie within the Ets-like recognition site (ns3831M5) reduced the neuron-specific enhancer activity of the 31 bp element by approximately
5 50% (Figure 7). From this it was concluded that a CCCTCCT motif, which partially overlaps the Ets-like recognition site within the 31 bp sequence, is absolutely required for neuron-specific enhancer activity.

This dissection of expression sequences using transgenic fish, exemplified in zebrafish and with GATA-2 as described above, provides a
10 system that allows the rapid and efficient identification of those cis-acting elements that play key roles in modulating the expression of developmentally regulated genes. Identification of these cis-acting elements is a useful step toward determining the genes that operate earlier than the gene under study in the specification of a developmental pathway
15 (since the identified distal regulatory elements interact with transcription factors which must be expressed for the regulatory elements to function).

Careful analysis of GATA-2 promoter activity in zebrafish embryos revealed three distinct tissue specific enhancer elements. These three elements appear to act independently to enhance gene expression
20 specifically in blood precursors, the EVL, or the CNS. Deletion of one or two of the elements will generate transgene constructs that can drive expression of a gene of interest in a specific tissue. Such constructs also allow study of the tissue-specific function of genes expressed in multiple tissues.

25 It has been shown that the developmental regulation of the mammalian HOX6 and GAP-43 promoter activities is conserved in zebrafish (Westerfield *et al.*, *Genes Dev* 6:591-8 (1992), Reinhard *et al.*, *Development* 120:1767-75 (1994)). If the same neuron-specific element identified in the zebrafish GATA-2 promoter is also shown to be required
30 for neuron-specific activity of the mouse promoter, one could specifically knockout expression of GATA-2 in the mouse CNS by targeting this cis-

element. This would allow one to determine precisely the role that GATA-2 plays in the CNS.

The neuron-specific enhancer element of GATA-2 has been precisely mapped and found to contain the core DNA consensus sequence
5 for binding by Ets-related transcription factors. Although Ets-related factors have been implicated in the regulation of expression of a number of neuron-specific genes (Chang and Thompson, *J. Biol Chem* 271:6467-75 (1996), Charron *et al.*, *J. Biol Chem* 270:30604-10 (1995)), another sequence, CCTCCT, present in this region of the zebrafish GATA-2
10 promoter was found to be required for expression in the CNS. This motif partially overlaps an inverted form of the core sequence of the Ets DNA binding recognition site. As has been shown for other genes, the activities of Ets family proteins often rely more on their ability to interact with other transcription factors than on specific binding to a cognate DNA
15 sequence (Crepieux *et al.*, *Crit Rev Oncog* 5:615-38 (1994)). It is possible that an independent factor that binds to the CCTCCT motif is required for neuron-specific activity of the GATA-2 promoter.

A number of growth factors are known to affect early embryonic expression of GATA-2. Noggin and activin, which both have dorsalizing
20 activity in *Xenopus* embryos, downregulate GATA-2 expression in dorsal mesoderm (Walmsley *et al.*, *Development* 120:2519-29 (1994)). BMP-4 activates GATA-2 expression in ventral mesoderm and is probably important to early blood progenitor proliferation (Maeno *et al.*, *Blood* 88:1965-72 (1996)). Growth factors that might affect expression of
25 GATA-2 in neurons are not known. However, both BMP-2 and BMP-6 can activate neuron-specific gene expression (Fann and Patterson, *J. Neurochem* 63:2074-9 (1994)). Consistent with studies on growth factors that upregulate or downregulate GATA-2 expression, GATA-2 promoter activity was excluded from the zebrafish dorsal shield. It has also been
30 discovered that lithium chloride treatment dorsalizes the injected embryos and dramatically reduces GATA-2 promoter activity as determined by GFP expression.

Although GATA-2 expression has not been observed in the EVL by *in situ* hybridization on whole embryos, this may be due to the conditions used. In mouse, embryonic mast cells present in the skin have only been detected by *in situ* hybridization performed on skin tissue sections (Jippo *et al.*, *Blood* 87:993-8 (1996)). Interestingly, expression of GATA-2 in mouse skin mast cells occurs only during a short period of embryogenesis, similar to what has been found for EVL cells in zebrafish. It is possible that the constructs used in this example may be missing elements that would specifically silence GATA-2 expression in the zebrafish EVL.

The method described above is generally applicable to the dissection of any developmentally regulated vertebrate promoter. Tissue specific and growth factor response elements can be rapidly identified in this manner. The fact that zebrafish typically produce hundreds of fertilized eggs per mating facilitates obtaining statistically significant results. While tissue culture systems have been useful for identifying many important transcription factors, transfection analysis in tissue culture cells cannot simulate the complex, rapidly changing microenvironment to which the promoter must respond during embryogenesis. Temporal and spatial analysis of promoter activity can be only poorly mimicked *in vitro*. The system described herein allows complete analysis of promoter activity in all tissues of a whole vertebrate.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: MEDICAL COLLEGE OF GEORGIA RESEARCH FOUNDATION
- (ii) TITLE OF INVENTION: TRANSGENIC FISH WITH TISSUE-SPECIFIC EXPRESSION
- (iii) NUMBER OF SEQUENCES: 27
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Patrea L. Pabst
 - (B) STREET: 2800 One Atlantic Center
1201 West Peachtree Street
 - (C) CITY: Atlanta
 - (D) STATE: GA
 - (E) COUNTRY: USA
 - (F) ZIP: 30309-3450
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: Patent In Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Pabst, Patrea L.
 - (B) REGISTRATION NUMBER: 31,284
 - (C) REFERENCE/DOCKET NUMBER: MCG100
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (404)-873-8794
 - (B) TELEFAX: (404)-873-8795

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: no
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCGGATCCTG CAAGTGTAGT ATTGAA

26

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: no
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AATGTATCAA TCATGGCAGA C

21

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: no

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
TGTATAGTTC ATCCATGCCA TGTG 24
- (2) INFORMATION FOR SEQ ID NO:4:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
ATGAACCTTT CTACTCAAGC T 21
- (2) INFORMATION FOR SEQ ID NO:5:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
GCTGCTTCCA CTTCCACTCA T 21
- (2) INFORMATION FOR SEQ ID NO:6:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
AGACACAGTC CAGGTGAGTC CAA 23
- (2) INFORMATION FOR SEQ ID NO:7:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
CTTTCGCCAC CTGGTATGTT GTG 23
- (2) INFORMATION FOR SEQ ID NO:8:
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AAAAAGAGGC TGGTATGTAA AA

22

- (2) INFORMATION FOR SEQ ID NO:9:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: no
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AAACTGCACA ATGTGAGTAT AC

22

- (2) INFORMATION FOR SEQ ID NO:10:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: no
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATTAAAACAG TTCGCAAGT C

21

- (2) INFORMATION FOR SEQ ID NO:11:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: no
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AATTTTACAG AGGCTCGTGA A

21

- (2) INFORMATION FOR SEQ ID NO:12:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: no
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCTGCATCAG ATTGTCAGCA AA

22

- (2) INFORMATION FOR SEQ ID NO:13:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: no
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CTTTTTGCAG GTCAACAGGC CT

22

- (2) INFORMATION FOR SEQ ID NO:14:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Arg His Ser Pro Val Arg Gln Val
 1 5

- (2) INFORMATION FOR SEQ ID NO:15:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Leu Ser Pro Pro Glu Ala Arg Glu
 1 5

- (2) INFORMATION FOR SEQ ID NO:16:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Lys Lys Arg Leu Ile Val Ser Lys
 1 5

- (2) INFORMATION FOR SEQ ID NO:17:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Lys Leu His Asn Val Asn Arg Pro
 1 5

- (2) INFORMATION FOR SEQ ID NO:18:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Trp Gly Ala Thr Ala Arg
 1 5

- (2) INFORMATION FOR SEQ ID NO:19:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 28 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: no
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

ATGGATCCTC AAGTGTCCGC GCTTAGAA

- (2) INFORMATION FOR SEQ ID NO:20:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: no
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TCTGCGCCGC TTTCTGCCCC CTCCTGCCCT CTT

33

- (2) INFORMATION FOR SEQ ID NO:21:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: no
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TCTGCGAAGC TTTCTGCCCC CTCCTGCCCT CTT

33

- (2) INFORMATION FOR SEQ ID NO:22:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: no
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TCTGCGCCGC TTTCTGAACC CTCCTGCCCT CTT

33

- (2) INFORMATION FOR SEQ ID NO:23:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: no
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCTGCGCCGC TTTCTGCCAA CTCCTGCCCT CTT

33

- (2) INFORMATION FOR SEQ ID NO:24:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: no
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TCTGCGCCGC TTTCTGCCCC AAACGGCCCT CTT

33

- (2) INFORMATION FOR SEQ ID NO:25:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: no
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCTGCGCCGC TTTCTGCCCC CTCCTGCCCT CTT

33

- (2) INFORMATION FOR SEQ ID NO:26:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5563 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: no
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GAATTCTAGT TCTAGGGTAA ACTATACAGT TTTTTTAATT AATAAAGTTG GTGGAGGTAA	60
ATGTCTTTAA TGAGTAAGTC ACTGAATCAT TTATTCATTT GATTGTGTTCA AACAGTTGAT	120
TCATTTAGAA ATTCATTAGA AATCAARCTG CAGTCTTTAT GAACGACCCG TTAAACCTTT	180
AGTTTATGTG ATTGGAATCA AAACCCCACT GTGTGTTAAT CAGATGAATG CTGAAAAGCA	240
CAGACAGGTT TTAATCCATC ATGCCATTCC TTCTAGAAAG GAAACATTAG TAATGGTTTT	300
AATTTTCAGC ATTTTAATAA CCACAAGCAC ATTTCTAATG CAATGAAATC ATATTTGCAA	360
ACCAAAACAG CTGATTCTTG AAATGGCCTA CACAGAGTCC AGACCTGAAT ATTATAGAGA	420
TGGTGCACTA TCACTTGAAA GAAAAATAAA CATTAATCTT AAATCTAAAG AACTTAAATC	480
TAAAGAAGCA CTATGAGAAA TGCTGAAAA GCCTGATTTT ACATAGCACA TTATTTAAAA	540
TGAAACCTCA GGgACAGTAT ACAGAACAGT TCAAATACAG TATACAGTAA ACAGAACAGG	600
TCAGGTCACA CCAAATACTG GCAAGCCATT TTATTCTGAA AATGTTTCAT TTAGATTAGA	660
ACAGAAGAAC TANAGAGACC NNNAAAGTTG GCTGAATATA AATAAATATA CCACTGCTTT	720
GACGGYTCTA GACTTTTGCA CAGTACTTAA ATGCAGTACT TAAAGTAATT CNTCATTTAG	780
ATGAGCTAAG TAAACTATGA GTTGTGAAAA AACACACCAT TGTGTGATGA GCAGTGAGGG	840
TGTCACTGTA GCTGTGAATT TGTTTCATGTA GTGCCATTAC TAGTTATACG ATCCCCAACC	900
TCCCACTCCA ATNTAGATAG CTTCTTATCA CAGTTCAGCA GCAGCGCACA CACACAGAAA	960
CACACACACA GCCACATCCN TCAAAANTGG TCTTTGGAGA CTTCTTTCTC TTTGACCGTT	1020
TAGTTTTCGT GAGCATAATT AAGTTACTCT ATACAATAAA ATGTGAGTAA ATGGACACCA	1080
TAGATGTCTA AATAAATAAA CACATAAATA AAAAGATGAC ACTTTCACAT AACACCATCA	1140
AACAGCTTCA TAAAATTATA TTATATAGAA TATTCTATAA TTATGTTGAT TTGTAACGCA	1200
CTGTAAAAAA AGGATTACTG CCTTAAATTG ATAATTGTT GAAGAAAATT TACTTTCCTG	1260
AACATTTATT GTATTAATAT ATTACAGTAC GCTCAATAAT ACATGTGAAA CTGCAGCTTC	1320

ATATTTTTTAA ATGTTTTTAAT GTATTTAATA TATATATATA TAATATTTAT ATATATATGT	1380
ATGCATGTAT GCATATTTTAT TCTGTTGAAA GGAGATTAGT TTTATTCAAC ACATTAGTTT	1440
TAATAACTCG TTTCTAATAA CTGATTTCTT TTATCTTTGT CATGATGACA GTAAATAATA	1500
TTTGACTAGA TATTTTTTCAA GACATTTCTA TACCACTTAA AGTGACATTT AAAGGCTTAA	1560
CTAGGTTAAT TAGGTTAAGT AAGCAGGTTA GGGTAATTGG GTAAGTTATT GTACAACAAT	1620
GGTTTGTCTT GTAGACTATT GAAAAAATG GCTTAAAGGG GCTAATAATT TTGTcCCTTA	1680
AAATGGTGTG TAAAAATGTA AACTGCTTTT ATTGTGGCTG AAAAAACAAA TAAGAATTTT	1740
TCCAGAAAAA AAAATATTAT CAGACACTGT GAAAATGTCC TTA CTCTGTT AACATAATT	1800
TGTGAAATAT GTAAAAAAGA ATAAAAAATT CaCATGGGGG GTGATAACTT CAACTACACA	1860
CACACACACA CACACACACA CACATTTTcAG tGACcAAAAT ATGTTGTRGG TTTNTKINTT	1920
CATTGATATA AAaTGTGCGA TGcCATTTTCM AAAATCCATA TATAGTTTAT GCAACATTAT	1980
ATTgGAMCCA AAATAAGTaA TATACAAAAT AAGTAGTATT ATCTTATCCA GTATATTTGA	2040
GTATTTATAT ATCGAAGTTT AGATTCTYTA TTTAACAATA TTTATGAATT ATATGTTTAA	2100
GTTCTAAAC AACACCTCAT GTAAATCAAT AACATGGTGC TTGGTACAGT ATGCTCAATA	2160
ATACATGAAA AACTGCAGCT TCATATTTAA AAATGTTATT GTATGCAATT ACATGTACAA	2220
TTACAAATAA CGTATGGTAA TGTATACAAA TATATATTTA GTAATAGAGG GTATAATATA	2280
TGTGATGCAC ATGCGAAAAA ATATATCACA CACACACGCA CGCACGCACA CACACACACA	2340
CACACACATT TATTTATGCA TATGTACACT ATAAAACCCA AAAAGTTAAA CTCAAACCAT	2400
TTAAGGAAAC TGATTGCAAC AAACCATTAA AGTTGAAAAA CGAATCCTAA TGAGTACTGT	2460
AAACTGAATN TATTTGAGTA AACGAAGCAA TTTGAGGACA GTAAAACCCA ATAAATGAAG	2520
AGAACTCAA CCAACTGAGC ACTGTAAAAC CTAACAAGTT AAGGCAACTC AAACCGTTTG	2580
AGGAAATCGA TATAAGAGTC CTGTGAACTG TATTTAATTA ACTCATTACT TCAAACTCT	2640
TTTCAAATTA GTAGAATTAA CATTcAGTAC ATTTTGAGTT ACTACACTCA TTTcATTGTA	2700
TAAAGTTGAC TGTTGGGTTT TACAGTGTAT CTTTTTATTA ATTTATATAA GAACATGTGT	2760
GGATAATATA AGTACATTTA TTAACATCAT TATATATGTG GCTTCAGCTT TATGCAAATG	2820
CTGAAAGTTA ACGAATTGAA ATCAATTAAG CATTTCAGTA ACATAACACG TATTGTAGGT	2880
TTTGTCTTCA TTGATATACA CATGCAATGC ATTTCAAGTC ATTTATAATT GATGCATTAT	2940
ATTGTATTGT ACCAATGTAA GTAATATATA ATATACTATA TTATATTATC CAGTATATTT	3000
GACTTTAAAA TATTAAAGTT TAGATTCCTA ATGTAACAAT ACATATATAA TATGTTAAGG	3060
TTCTAGAATG GAACCTTATG TAAATCAAWA ACCTGGCGCT TGGTGAAGGA TTTGCTTCTC	3120
TGRATCTCa CCCAGTTTCC CTGAAAATTA TAAATGCACA ATGGTGGARG GAAGTTGAAA	3180
GTGtTTTGCC TGTCAAATGA RARTGACAGT CTTAGTCctG TGCTCCGgCA GSCCGTTCTG	3240
CGTCCGTATC TCTCACCATG ATTGCAGCAT TKGAGTTTAT TTGCATTACT GTTCTTTGCT	3300
GAGCTGCACC AgGGGAAAAG TGCTTTTGCA TTTTCATTcG CTTTGTTcAC AGTCACCGTT	3360

TCCATCCCAA	GTGCTCTTTG	TTAACACTTT	GCACGCCATT	TTAATTGCCA	AATGTATTAG	3420
GCCACAGCAT	ATGCTTAATT	CTTTTCAACA	ATGAACTTTT	ATTAATGATG	TGCTTGAATC	3480
ATAGATACTA	TAAGTTTATG	GTTGTTGTAA	AATTARGTTT	CTCTGGCTGT	CTGTGGGATT	3540
TTCCCAGCGC	TGTTGGATTT	GCGTCTTTAT	CTATATTTAT	AAGTGAAGCC	ATTTTATATA	3600
ATCTCTGACA	GTATTTTATT	TAGATTAGAA	ATTAAATACT	AGTGTTTTTT	GTCTTGTTTC	3660
TATAGTATTA	TTACTATTTT	TTTGCATTAA	TTTACAGAAG	ATGCCTGATA	AACTGAATTT	3720
AGTATAATAA	TTTAAATACC	AAAACATCAT	TAGGTACATT	TAAAATACCA	ATCATGCAAA	3780
AAAATAACCC	TTTGA CTGCA	CATTTACCCA	ATGGGTGTCC	ATTTTTGACT	TTTTAAATAA	3840
TGGTTTACAC	ACACATCATT	GCTGTTTTAC	AAAAAATCA	AACATAATTC	TTTTGCACGA	3900
CTACTCTGAA	TTTTGGTTTC	ATTCATTTTC	TTTTTGGCTA	AGTCTGTTTA	TTAATATGGA	3960
GTCGCCACAG	CGGAATGAAT	CGCCAACTTA	TTTAGCATAT	GTTTCACACA	GTGGATGCCC	4020
TTCCAGCTGC	AAACCATCAC	TGGGAAACAT	CCATACACTA	TGGgACAATT	TAGCCTACCC	4080
AATTCATCTG	AACTGCATGT	CTTTCAGGg	AAACCCACAC	AAACACgGGG	GAGAACATGT	4140
TTGGTTTAAT	TGTAAAAAAA	CAACCAGAAA	GCATAATAAA	TGAGAATCTC	AAATATTTTT	4200
ACCGCATACT	TCAAAAATAA	AGATGATTTA	GTATTAAAAA	ATGTTTTATT	TTGAATATtG	4260
CTTTTAAATA	AATTGGSCTT	ACaCTTAGTA	TATGTAtTAA	TTCCAGTACT	TTTACCATAA	4320
ACCGACATAT	CMACCATTtG	G TAGAGGTtG	ATAtTTTAGA	AATGACgARA	WGTGTTGAAA	4380
AAAAtGCATC	gAGTGTGTAg	CAACATTAGG	ARTTAAGTAT	TGCAAtGCAA	AAaTtGTaAg	4440
TWAATCAATt	AGGGACTaAT	TAWTCGTCAA	TTTAAATTGT	TATAATTTGc	TACTTTTTCT	4500
CAAACCACTA	GGTTTCACTG	ATTATTCAGC	AAAATGTTAT	TCATCATTTT	CAATTTTATA	4560
TATTTTAACA	TGAGCAGCAT	TTTTACTTTA	ATATATACTG	CACAAAAAAT	AGTTACATTG	4620
TGTTTTTAAG	CGTTTCCTTT	ATTTATTTAT	TTTTTTGAGC	AGTATATTTT	TAAAAAGTGA	4680
GAATAAATAT	G TAGCTTTAG	TTTACATAA	CCATATGATG	CACTTAACGA	TGATGAAACA	4740
TTTCATTCAT	ATTTGGGGCA	TTTTATTTTT	ACTTATTTTT	TTTGAAAAAA	TGGACACTAA	4800
CTGTGGTTTT	AATATGATTT	CTATGTAAAT	AAAATGACTT	TTGGACATTT	AATTTGATGT	4860
ACACTGTAAA	AAAAATCCAA	CCTTAAATTT	TAAGTTAAAT	CAAGTTAACC	TTATCAGTAC	4920
ATTGAACTTA	AATTATGTTA	AACTGACATA	AACTGAATG	AATAACTTAT	AAAATTAAGT	4980
TAGAACACCA	TAGATTAATG	TTACAATGAA	CTAAAACTG	TCATGACTAA	TTGTTTCATAT	5040
TTATATTTTT	ACAGTGTAGA	TGTGGAACAT	CCAGTCTTTG	TYTATAAGGT	CATATAGGCT	5100
AAAATYTAAT	AAAACATTTA	AATAGGAATT	AAAATTTTTG	TTTCTTAATA	TTTTTATTGT	5160
AATTTCTTAA	CATTTACTCA	GTGAAACTAA	TTTCAGTTTT	GATTCTTTCA	CTATAATATG	5220
TGTATATATG	TGTATTATAA	AAATAATTTG	TGTTCAAAAT	AAAATAAAAA	AATTTGCACA	5280
ATCCTCCACT	ATTCATTTGA	ACTGAACTCA	CATGCTGTGT	CAGCTAGAGA	TCTGCCATAT	5340
AATATTCAAA	ATGGAAAGCG	TGGCCACCCG	TATGGTAGGA	GTGTCCAAAA	AAAAGTACCC	5400

CAACCCACC CATTGGTGCC CTACAATTC AAATGAACCT ACTAGTCCC AAAGACTGAA 5460
 GGAGATAAGC AAGCAAACAG GCGGCTAGTT CACTCCATGA TCTGAGaATC TCCTGRYACT 5520
 GATAAACGAC ATCTTCAATA CTACACTTGC AGGATCCACT AGT 5563

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4811 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATATTTTGGG TTATGGCTAA AATAATTAAT GTCTAAAACG GGATTACGCG TTTTTCGTAA 60
 AGCTCAAAGA CGCATGTGCC AAAAATAGCC TTTTATTAAA TTGTTTGGTT ATTAAAATAT 120
 TATTCAACTT ATTTTACATC CATGGAAAGA GACATGGCCT CTTCTATTTG ACCTGCATGT 180
 GTTAAACGA AATGCCAAAA TAAAGAAAAA AATGTAATTC AACATGTAAG GCTATTCAAA 240
 AACAAACAC AGGTACAAAA CATATCTTTG TTAATGAAAC TAATTTACAG TTTGTTTATT 300
 AAAACACACT ATAAATGCCA TAGAACATTT TGGAGATGCA TGC GTTATAC ATTGCGTGAT 360
 TTAACAGATC AATTAAAGTC GTATTTTTCG CCAGCATTTT AATGGGCATA ACGACTTAAT 420
 GTTTTCCTCT AGAATGATTA CAAATGTGAA AGCGAATGTG ATGTGATTGA GTTGAAGAAT 480
 TAGTTTTTTT TGGAATGCCC CAAGGACGCA TGCATTAGCC CACCTGTGCT GTTTATTTAA 540
 ATCATTGACT CCAAGAGCTG TCAGCCACAA AAGGAGGGCG GGCGCGCTGT CATCACCCAT 600
 CAGATTTATG ACTGCCACAC AATCATTTTC CGACTAAACT AACGCCATCA TCACTCAGAA 660
 CAAGAACTTC ATGAGTCGCA CAAGACAAGT TATAATAAAT GCATTACAGC GAATGCATGC 720
 ACAAACGCGA GAACCACTTT TGCTGCAAAA TAATGTGGAT TGTTGGTTGA AATGAAAAC 780
 GGGTGAGATG CTTTTCTTTC AATCCCTGTT ATCCATGCTT CAGCAGAGGA CAGGAGGCTT 840
 GTGACTTTGC CTGTGCCTGT GTCTGCCCCC GAGTGCCCTG TCACAATCTA ATTACCCGTG 900
 AGTAAAGGAC AATACCGCTT CAGCTGGTCT GTGTCAATCC CCCTATATCC CAGTGCCTGC 960
 TTATTTTTCAC AAACCCTTCT GCGCCGCTTT CTGCCCCCTC CTGCCCTCTT TTAACCCAC 1020
 GGAGAATGAT AAATGCGCGG TGAGGGAACG AACGGGCAAA GCCATTTTAC GGCACCTGTT 1080
 AATTAAGGGA ATGATTGCCT CCATTTTTCG CTGAGCTCGT TTCCAGCGTG CTCCATTATT 1140
 TGTGATGCGA TTAATTGAAA GCGAATGTGA CATCACAACG AACGTGATGT CATTGTGCGC 1200
 GTCACACAGT AGAACGACAG AGTTACATAA GAAATAAAGT CTGCATGCAT ACATTTATGC 1260
 ATGGCGTTTT AAAGAAGAGC GCACACTGGG TTAGAGTCCT CGGTGGGGTC AGCCACTTCG 1320
 GTAACACCCC AAGCATTCAA TGCTAAGCCC TTAAAAGGAC AGCGTCTTTT GTTCTAACAT 1380
 CGAGAGCACC GGGATTACCA CAGGTATTTA GTTCAGGTAT TCTCTAAGAA TATTTAGCCC 1440
 TAGGTGAGCT GAACCAAGAG CAGTCATTAG CGCTAAACT GGCTCTGATG GGAAGGGCTA 1500

ACACACACAC ACACACACAC ACACACACAC ACACACACAT TATAATAAAT GTAATGTCAT	1560
GTTTACAACA ACTCCGGCAG TGATGCTGCA TATTGGCGGC GTACATACAC TAAATGTTTT	1620
AATGTAGTCT GTAAGACTAG AGAATCAGAA ATTAATTTAC ACAGAAATTA CAAAAATAAA	1680
TACATGTTTA AATAGTTAAT AAACATAATT CAAATATGTA ATGTATTATC GTGTATTTTA	1740
ACATTAATGG ATGAGGTGGT TCAAATGCAT TTTGCACAAA ATAAAATCGA AGCAGCTTCA	1800
AATCGTAAAG ATAATAGTCG GTAGCATTGA ATCTGCTTTA ACATTTACTT TTAGCGAAGG	1860
CTACTTTATT AAGGAAGCTC ATATTAACTC CCAATGAATG TCTGCTATTG CACCTTTTTG	1920
AGGTGTAGAC TGTGTAAAAT GCATCACTGC ACAGCAAAAT CAAGCGTCAT ATTATCCTGT	1980
ACATTCTAAT TTGTTGGCTT CAGGCTGCCA GGGCTCTTTG TGCTGTGTAG GGCCCCCTGGC	2040
CAGATTCCAG TGTGTTAAAA AGGGATTTAC GCATCTGATA TTGTCACACA ATAAGGACAA	2100
ATAGCCCGTT TGAGCATCTT TATACAACCA ACGCTGACAG AGGTTCTGCG GTTTAAGTGC	2160
TTAGTGTTGC ATTTGTGCTT AAATTGATTG TTTGGTGTTC AACCCTCACT GGAAAAAAT	2220
CTTTTGATGC AAATGGGTGC GTTTAGATAA AAAGAAGCAA AGCCTAGAAC TAAAGCCTAG	2280
AATTTATATT GCACTGTAGA TGTGGATGGT TATGGGAAAG TTTTTTGAGA TACTGTGGGG	2340
CGAGTCACGG CGTCAGAGTG GCGGCCGGTA GGGGCTCTAA ACTCGCGCTC CAATTATTGC	2400
CTGTCACTCA TCATCGCTTT AGATTAGAGC ATGCGGATTA AAACCTCATGC CTTTAAATAA	2460
TAACAACAGC GTCAATATTA TCAAAAAGAC ACATCACGCT TATTTAAAT CTACGAAATG	2520
TGTTAAAGCA TAATTTGTAC TACTGGTTGA TTGTTGTAGA CCTGAAATCC TGTCAGATAG	2580
AAATGAACTA CCCGGACCAC TGGTAGTTAA GTCTCTCTTG TGTTATCTTT GATTGATCCA	2640
ACCAGACAAG CTAGTTAAAT TAATAATTTA TAAGCGCAA GCGTTGGTAC AAGCAGTTAG	2700
AGGGAGAAAG GTGAGAAGAA GCAATACAAA GTAGCTAAAT TCACAATGCA TTACATTGTC	2760
CATTTTAGAA ATGAAACACG AGGATTTAAT GTTAAATGAA TACAGAGTAG CTATAATCAG	2820
CAATACAAAG TAGCTAAATT CAGCAATACA AAGTAGCTAA ATTCAGCAAT ACAAAGTAGC	2880
TATATTCAGC AATACAAAGT AGCTAAATTC AGCAATACAA AGTAGCTATA TTCAGCAATA	2940
CAAAGTAGCT ATATTCAGCA ATACAAAGTA GCTAAATTCA GCAATACAAC GTAGCTATAC	3000
TTTGTAGCTA TACACTGTAT CCATTTTAGA AATGCACACG ATGATTTTCT GTTAAAAATC	3060
ACTGCTCATT TGAATTAGAT TATTTGAATT GGAGCTTACA TTGCATGTAA TTAGTAAGCA	3120
AATTCGGCTT AACAAATTTG AAACGCGTTT TTTTTCTCG ACTAAATTAA TTAAGAAAAT	3180
GTATTATTGA TGGGTGCAAA CAGTAACAAT TTATTAAACC CTCTATGCAA ATGAGGTGTT	3240
CAGCTGACTA ACCTGCATCC ACAGTTTATC TAAACGCTTA TCAAACATAAT TGGCGACGTT	3300
CTGTCTTTCT GCCTGCGGTG GCGGAGCCTG CTGCTTGTTT TGCCACGAGA TAATTGTACG	3360
CAAGAATCAA CGAAGCTGCC CTAATGGCCA CCAATTGGCT TTATTTGGAC CTGCCCATGC	3420
GACCTGTCTG CACCTCCAAG AGACGGGCTC GCTATTAATA TGTAAGTGGA CGTTTGATCG	3480
CTTGAAACGG CATACAAAGA CAGTGTTTTC ACAAGAAGAA TGTGGTGACA ACTCATTTAA	3540

AACTATTAGA	CGCGCAAGAA	CAATAGCCCC	CAATTTAGAG	ACCATAAAAT	ACTCCTCCCC	3600
AATTAATGCC	TGAGGTGCTA	GGAGTTGAGT	TTGCTTGCAT	TAGGCACATA	TCTCATGTGA	3660
CACTTCAGTG	TTACAGGTTT	TGTTGTTTTA	AGCTAATGTT	AATGGTCAGG	GAACAGCTCG	3720
TAATCACAAT	ATATATTTAA	AACAAATGAT	TATTATGAAT	GCAATAGGCC	AAATCGATAT	3780
TCATTAATAG	AATAGAGGCA	TTTAATACA	TTTCTGCACA	ATTAAAAATT	AAATATAATC	3840
CTGCAAGTCT	ATAATTATAT	TATTCACATC	ATTTAATGTC	CTAAAAATAA	ATTTAAAAAA	3900
TAGCATTAGG	CTGCAACTTA	GATTTTAGGC	TTTTCTGTTA	GCACTTGAGT	AAAAAGACAT	3960
CATTACACAC	CATCAACGTG	AAGCTCTAAA	AAGGGTAAAA	AGATCTCAAT	AAATTGCTGC	4020
GCTGAATGAT	GAGTCTCTCA	GCTCTCTGGA	TGTGGAGCAG	TAGGCCGACA	GTCGCCGTGG	4080
CATTTCCGAA	AGCATGCTGT	CCGAGCCAAT	GGCAGTCAGC	GCGCTCTGCT	ATTGGTTCCC	4140
AGGGCGCTCA	CTGCCAGCTC	GTGTCCCCGC	CCATGTTCTG	AAGATATGGA	ATCTACTGGC	4200
GCCAGTTCCG	ACAGTACACA	GGCACAATTC	ATTAATGAGA	CTTCTCTCCG	CTTTAGACAG	4260
ACGCAGAGTT	TTAGGGAGAC	TTAACAATC	GGGCTGTGGA	CAATTTAAAC	CAGTGGCGAA	4320
TTACGAACGT	CAACAGGCAT	CTTGAGGATT	AACATTCTTT	GCGCAGGACT	AACACGGGAA	4380
AAATAAACGC	AGGATTGGAG	TGCTGAAATG	CAACTTTGCG	CCGTGAGTAC	TTCCCGATAG	4440
TTATTTGAAA	TTGCGAGCAT	TTAATTGAGC	GATTTAATTG	ATTGACTACA	AAAGTTAGCC	4500
TACTTATATT	AACTGAGGCG	TCGTCGTGTG	AATTAAGATC	TGTCTTGAC	TGTGTTTAAC	4560
GTCAACACTG	AGATGCTTCT	ATCTGTTATT	CTCTTACAGG	TGTCCCTGGC	CACCCTTGAA	4620
TGCAAAGAAG	CAGGACCTCT	ACACTCCTTC	AAAAATAAAA	GCATGCTCAG	AAAGTAAACA	4680
GAGCATCGCC	ACCTGAAGCA	TTAAGCTAAC	GACAGATATT	TTAATAATCT	AACGGACTAT	4740
AGTGGTGCTT	TCGGGTCTGT	AGTGTCAGT	AACTTTTCC	AAGCATTTTC	TAAGCGCGGA	4800
CACTTGAGAT	G					4811

CLAIMS

1. A transgenic fish the cells of which contain an exogenous construct,

wherein the construct comprises homologous expression sequences operably linked to a sequence encoding an expression product, wherein the expression product is expressed only in specific cell lineages.
2. The transgenic fish of claim 1 wherein the expression sequences and the sequence encoding the expression product are not operably linked in nature.
3. The transgenic fish of claim 1 wherein the expression product is heterologous.
4. The transgenic fish of claim 3 wherein the expression product is a reporter protein.
5. The transgenic fish of claim 4 wherein the reporter protein is selected from the group consisting of β -galactosidase, chloramphenicol acetyltransferase, and green fluorescent protein.
6. The transgenic fish of claim 5 wherein the reporter protein is green fluorescent protein.
7. The transgenic fish of claim 1 wherein the fish is selected from the group consisting of zebrafish, medaka, trout, salmon, carp, tilapia, goldfish, loach, and catfish.
8. The transgenic fish of claim 7 wherein the fish is zebrafish.
9. The transgenic fish of claim 1 wherein the expression product is expressed only in cells selected from the group consisting of blood cells, nerve cells, and skin cells.
10. The transgenic fish of claim 9 wherein the expression product is expressed only in blood cells.
11. The transgenic fish of claim 10 wherein the expression product is expressed only in erythroid progenitor cells.
12. The transgenic fish of claim 9 wherein the expression product is expressed only in neurons.

13. The transgenic fish of claim 1 wherein the expression sequences are selected from the group consisting of GATA-1 expression sequences and GATA-2 expression sequences.

14. The transgenic fish of claim 13 wherein the expression sequences comprise GATA-1 expression sequences.

15. The transgenic fish of claim 13 wherein the expression sequences comprise GATA-2 expression sequences.

16. The transgenic fish of claim 15 wherein the expression sequences comprise the GATA-2 promoter operably linked to the neuron-specific enhancer of GATA-2.

17. The transgenic fish of claim 15 wherein the expression sequences comprise the GATA-2 promoter operably linked to the blood-specific enhancer of GATA-2.

18. The transgenic fish of claim 15 wherein the expression sequences comprise the GATA-2 promoter operably linked to the skin-specific enhancer of GATA-2.

19. The transgenic fish of claim 1 wherein the transgenic fish developed from, or is the progeny of a transgenic fish developed from, an embryonic cell into which the construct was introduced.

20. The transgenic fish of claim 1 wherein the expression product is expressed only in predetermined cell lineages.

21. The transgenic fish of claim 1 wherein the exogenous construct is genetically linked to an identified mutant gene.

22. The transgenic fish of claim 1 wherein the expression sequences comprise a homologous promoter operably linked to a homologous enhancer.

23. The transgenic fish of claim 22 wherein the expression sequences further comprise homologous 5' untranslated sequences operably linked to the promoter and the sequence encoding the expression product.

24. The transgenic fish of claim 1 wherein the construct further comprises (a) intron sequences operably linked to the sequence encoding the expression product, (b) a polyadenylation signal operably linked to the sequence encoding the expression product, or both.

25. Cells isolated from the transgenic fish of claim 1 wherein the cells express the expression product.

26. A method of making transgenic fish, the method comprising

(a) introducing an exogenous construct into an embryonic cell of a first fish, wherein the construct comprises homologous expression sequences operably linked to a sequence encoding an expression product, and

(b) allowing the egg cell or embryonic cells to develop into a second fish, wherein the expression product is expressed only in specific cell lineages of the second fish.

27. The method of claim 26 wherein the expression product is expressed only in predetermined cell lineages.

28. The method of claim 26 wherein the method further comprises producing progeny of the second fish.

29. The method of claim 26 wherein the expression sequences and the sequence encoding the expression product are not operably linked in nature.

30. The method of claim 26, wherein the expression sequences are expression sequences of a fish gene, wherein the method further comprises

(c) exposing the second fish or progeny of the second fish to a test compound,

(d) detecting the expression product in the fish exposed to the test compound, and

(e) comparing the pattern of expression of the expression product in the fish exposed to the test compound with the pattern of expression of the expression product in the second fish or progeny of the second fish not exposed to the test compound,

wherein if the pattern of expression of the expression product in the fish exposed to the test compound differs from the pattern of expression in the fish not exposed to the test compound, then the test compound affects expression of the fish gene.

31. The method of claim 26, wherein the expression sequences are expression sequences of a fish gene, wherein the method further comprises

(c) detecting the expression product in the second fish or progeny of the second fish,

wherein the pattern of expression of the expression product in the second fish or progeny of the second fish identifies the pattern of expression of the fish gene.

32. The method of claim 26, wherein the expression sequences are expression sequences of a fish gene, wherein the method further comprises

(c) crossing the second fish or progeny of the second fish to a third fish having an identified mutant gene to produce a fourth fish having both the exogenous construct and the identified mutation,

(d) detecting the expression product in the fourth fish or progeny of the fourth fish, and

(e) comparing the pattern of expression of the expression product in the fourth fish or the progeny of the fourth fish with the pattern of expression of the expression product in the second fish,

wherein if the pattern of expression of the expression product in the fourth fish or progeny of the fourth fish differs from the pattern of expression in the second fish, then the mutant gene affects expression of the fish gene.

33. The method of claim 26, wherein the method further comprises

(c) crossing the second fish or progeny of the second fish to a third fish having an identified mutant gene, wherein the exogenous construct and the mutant gene map to the same region of the genome, to produce a fourth fish having both the exogenous construct and the mutant gene, and

(d) crossing the fourth fish to a fifth fish, wherein the fifth fish has neither the exogenous construct nor the mutant gene, to produce a sixth fish, wherein the sixth fish has both the exogenous construct and the mutant gene,

wherein the mutant gene is marked by the exogenous construct in the sixth fish.

34. The method of claim 33, wherein the method further comprises

(e) crossing the sixth fish, or a progeny of the sixth fish, with a seventh fish, and

(f) identifying progeny fish expressing the expression product, wherein fish expressing the expression product have the mutant gene.

35. The method of claim 26, wherein the construct comprises a homologous promoter operably linked to a sequence encoding an expression product, wherein the promoter is not operably linked to an enhancer, wherein the method further comprises

(c) detecting the expression product in the second fish or progeny of the second fish,

wherein if the expression product is detected, then the exogenous construct is operably linked to an enhancer.

36. The method of claim 35 further comprising

(d) isolating the enhancer from the second fish or progeny of the second fish.

37. The method of claim 35 further comprising

(d) determining the pattern of expression of the expression product in the second fish or progeny of the second fish,

wherein the pattern of expression of the expression product in the second fish or progeny of the second fish identifies the pattern of expression of the enhancer.

38. A method of identifying regulatory elements in sequences upstream of a gene of interest, the method comprising

(a) introducing members of a set of exogenous constructs into separate embryonic cells, wherein each member of the set of constructs comprises a sequence encoding an expression product operably linked to upstream sequences of a homologous gene of interest, wherein the different members of the set have different regions of the upstream sequences deleted,

(b) allowing the embryonic cells to develop into fish,

(c) detecting the expression product in the fish or progeny of the fish,

(d) determining which regions of the upstream sequences are needed for expression of the expression product.

39. The method of claim 38 wherein determining which regions of the upstream sequences are needed for expression is accomplished by comparing

the expression of the expression product in fish into which different members of the set of exogenous constructs has been introduced,

wherein if the expression product is detected in cells of interest in a fish, then the exogenous construct introduced into that fish includes a regulatory element for expression in the cells of interest,

wherein if the expression product is not detected in cells of interest in a fish, then the exogenous construct introduced into that fish does not include a regulatory element for expression in the cells of interest.

40. A nucleic acid construct comprising expression sequences derived from fish operably linked to a sequence encoding an expression product, wherein the expression sequences comprise a promoter operably linked to an enhancer, wherein the expression product is expressed only in specific cell lineages.

Figure 1A

IVS-1
 AGACACAGTCCAG (GTGAGTCCAA....1.6 kb....ATTAAAACAG) TTCGCCAAGTC
 R H S P V R Q V

IVS-2
 CTTTCGCCACCTG (GTATGTTGTG....0.07 kb....AATTTTACAG) AGGCTCGTGAA
 L S P P E A R E

IVS-3
 AAAAAGAGGCTG (GTATGTAAAA....1.7 kb....CCTGCATCAG) ATTGTCAGCAAA
 K K R L I V S K

IVS-4
 AAAGTGCACAAT (GTGAGTATAC....0.08 kb....CTTTTGCAG) GTCAACAGGCCT
 K L H N V N R P

Figure 1B

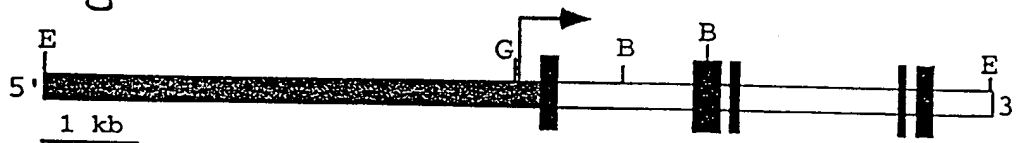


Figure 2

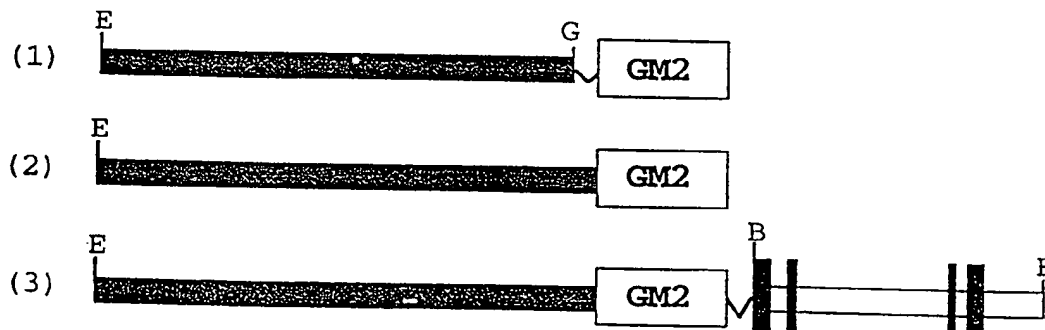


Figure 3

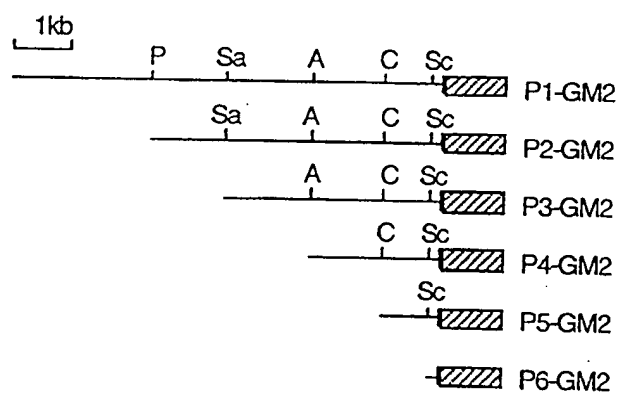
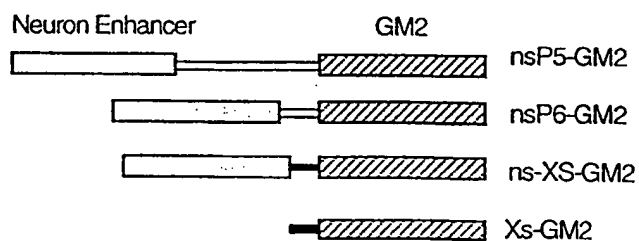


Figure 4



3/4

Figure 6

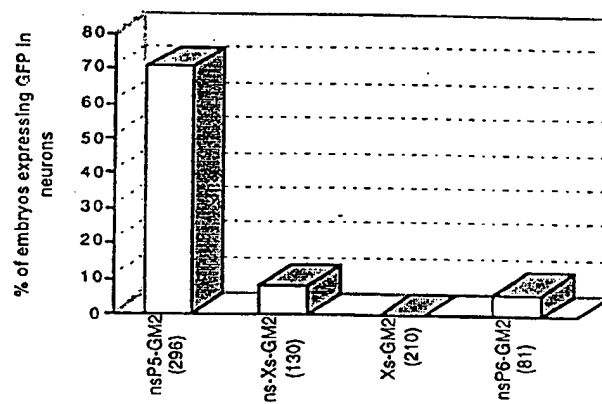
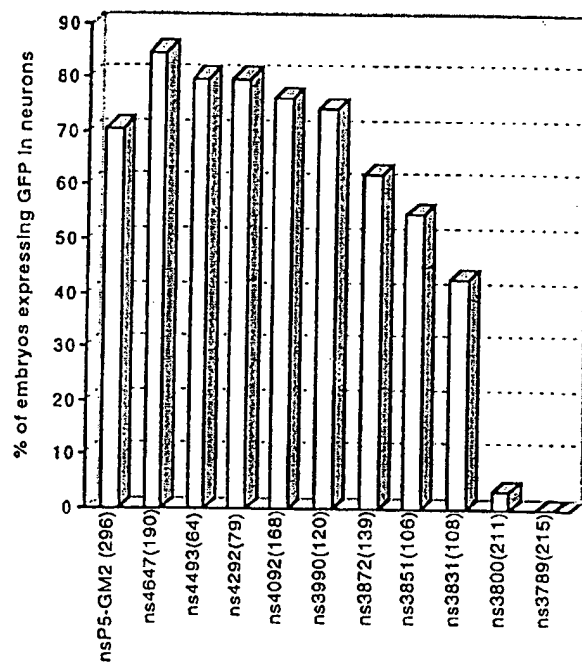
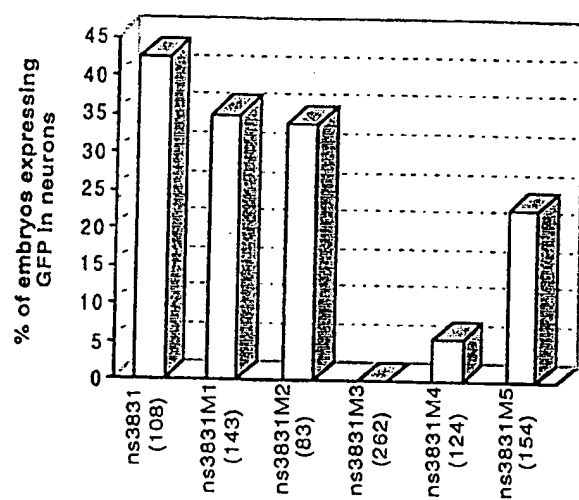


Figure 5

Figure 7



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☒ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☒ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.